



La régulation des éléments transposables par la voie des piARN : Les différences entre lignées germinales mâles et femelles et leurs conséquences sur la dynamique de transposition

Bastien Saint Leandre

► To cite this version:

Bastien Saint Leandre. La régulation des éléments transposables par la voie des piARN : Les différences entre lignées germinales mâles et femelles et leurs conséquences sur la dynamique de transposition. Biologie moléculaire. Université Paris-Saclay, 2016. Français. NNT : 2016SACLS084 . tel-01305511

HAL Id: tel-01305511

<https://theses.hal.science/tel-01305511>

Submitted on 21 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° NNT :
2016SACLS084

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS -SACLAY
PREPAREE A
L'UNIVERSITE PARIS SUD

ED n°577 SDSV | Structure et dynamique des systèmes vivants
Spécialité du doctorat : Sciences de la Vie et de la Santé

Par

Bastien Saint Léandre

**La régulation des éléments transposables par la voie des *piARN*:
Les différences entre lignées germinales mâles et femelles et
leurs conséquences sur la dynamique de transposition**

Thèse présentée et soutenue à Gif-sur-Yvette le: **24/02/2016**

Composition du jury :

Directeur de thèse :
Co-directeur de thèse :

CAPY Pierre
HUA-VAN Aurélie

Professeur, Université Paris-Saclay
Maitre de Conférence, Université Paris-Saclay

Rapporteurs :

RONSSERAY Stéphane
COLOT Vincent

DR CNRS au LBD, Université Pierre et Marie Curie Paris
DR CNRS à IBENS, Paris

Examineurs :

DREZEN Jean-Michel

DR CNRS à IRBI, Université François-Rabelais de Tours

Président du Jury:

DE VIENNE Dominique

Professeur Université, INRA-Moulon

SOMMAIRE

INTRODUCTION.....	1
I- ASPECTS HISTORIQUES.....	3
1- Barbara McClintock : la découverte des éléments mobiles	3
2- Les plasmides et les phages : les épisomes	6
3- La convergence des arguments : la notion étendue d'éléments génétiques mobiles	8
4- Les éléments transposables, des facteurs majeurs de l'évolution	10
5- La contribution des modèles : «du neutralisme au sélectionnisme»	13
6- Les éléments égoïstes auto-régulés	14
7- Base expérimentale et modèles pour la régulation.....	15
A - Inhibition somatique.....	15
B - L'empoisonnement par les copies non-autonomes.....	15
C - Les prémices de la régulation par ARN interférences.....	16
II- DIVERSITE DES ET DANS LE VIVANT	19
1- L'architecture des génomes	19
2- ET et procaryotes	22
3- Diversité des ET chez les eucaryotes	24
A - Quelques exemples extrêmes	24
B - Les ET chez les Arthropodes.....	26
C - Les ET chez les vertébrés.....	26
III- SEXE ET ET: L'ART DE FAÇONNER LES GENOMES	34
1- Origine des chromosomes sexuels	34
2- Conséquences évolutives de la différenciation des chromosomes sexuels	36
A - L'accumulation des ET chez les chromosomes sexuels anciens	36
B - L'accumulation des ET chez les chromosomes sexuels émergents	37
C - Les ET présents dans la région du déterminisme sexuel (SD).....	37
3- Dégénérescence et dégradation	38
4- Les ET dans la lignée germinale mâle	40
A - La surexpression de l'activité des ET dans les testicules.....	40
B - La surexpression dans les testicules est favorisée par la sélection sexuelle.....	41
5- Les adaptations liées au sexe médiées par les ET	43
IV- REGULATION DES ET DANS LA LIGNEE GERMINALE	45
1- Le maintien de l'intégrité du génome: les <i>pi</i> ARN dans la lignée germinale.....	45
2- Les mécanismes de biogenèse des <i>pi</i> ARN.....	47
A- La biogenèse primaire.....	48
B- La boucle d'amplification ping-pong	50
3- Les effecteurs protéiques de la voie des <i>pi</i> ARN	51
A- Les protéines à domaine <i>TUDOR</i>	51
B- Une large diversité d'interactants protéiques.....	52
4- Régionalisation cellulaire au cours de la biogenèse des <i>pi</i> ARN.....	53
A - Le complexe nuage	53
B - Les granules Yb des cellules somatiques de l'ovaire.....	54
5- La co-évolution des mécanismes de régulation et le génome hôte	55
A - Un modèle analogue aux interactions hôtes-pathogènes ?.....	55
B - La divergence des protéines associées à la régulation par les <i>pi</i> ARN : une diversification extrêmement rapide	56
PROBLEMATIQUE.....	61

RESULTATS.....	63
ARTICLE N°1	65
La voie des <i>piARN</i> limite l'expansion de <i>mos1</i> dans la lignée germinale	
ARTICLE N°2	105
Transcriptional polymorphism of <i>piRNA</i> regulatory genes underlies the <i>mariner</i> activity in <i>D. simulans</i> testes	
ARTICLE N°3	139
<i>piRNA</i> and transposon dynamics in <i>Drosophila</i> : a female story	
DISCUSSIONS.....	175
I- MODELE DES DYNAMIQUES D'ET ET DE LA REPONSE GENOMIQUE ASSOCIEE	177
1- Un équilibre dynamique entre allèles répresseurs et copies actives.	177
2- Le polymorphisme d'expression des gènes de la voie des <i>piARN</i> affecte l'activité de <i>mos1</i> dans les populations colonisatrices.	178
3- La réponse du génome par les <i>piARN</i> fait suite à d'importantes vagues d'amplification.....	179
4- La régulation par les <i>piARN</i> chez la drosophile, une spécificité de la lignée germinale femelle inopérante chez les mâles	180
II- LE CYCLE DE VIE DES ET ET LEUR REGULATION AU NIVEAU DES CLUSTERS DE <i>PIARN</i>	181
1- L'élément <i>mos1</i> est dans la phase stationnaire du cycle d'invasion des ET	181
2- La dynamique des ET et leur régulation par les clusters de <i>piARN</i>	183
3- Le délai de réponse de la régulation par les <i>piARN</i>	188
4- Les clusters de <i>piARN</i> : le cimetière des ET	190
III- LES CONSEQUENCES EVOLUTIVES DE L'ACTIVITE DES ET DANS LE GENOME	194
1- ET et pathogènes : des conséquences similaires pour l'évolution des génomes ?	194
2- L'hétérochromatine, une source de conflit favorisant la diversification des gènes associés à la régulation des ET.....	196
3- La divergence entre populations naturelles : une question peu étudiée	198
IV- LA GUERRE DES SEXES.....	199
1- Dimorphisme sexuel et sélection sexuelle	199
2- Les gènes « sexe-biaisés » sont très répandus dans la lignée germinale et évoluent rapidement.....	201
3- Le sexe : une force motrice dans l'évolution des gènes effecteurs de la voie des <i>piARN</i>	204
V- LES MODELES « MALE DRIVEN EVOLUTION » ET « OUT OF TESTES »	205
1- L'évolution facilitée par la lignée germinale mâle (« male driven » evolution)	205
A - Le nombre de réplifications dans la lignée germinale mâle	205
B - Le remodelage hétérochromatinien au cours de la spermatogenèse	206
2- Les testicules, une niche favorisant l'émergence de nouvelles fonctions (« <i>out of testes hypothesis</i> »)	207
CONCLUSION	211
BIBLIOGRAPHIE.....	215

Introduction

Notes :

Pour des raisons d'homogénéité dans l'écriture de ce manuscrit :

- les gènes et allèles seront écrits en *minuscules italique* (ex : *ago3*, *mos1*)
- les protéines seront écrites avec la première lettre en majuscule (ex : Ago3, Aub)
- les familles de protéines ou clades seront en MAJUSCULE (ex : famille des protéines PIWI)
- les domaines de protéines seront écrits en *MAJUSCULES ITALIQUES* (ex : domaine *TUDOR*)

I- Aspects historiques

« I concluded that regulation of gene expression was associated with an event occurring at a mitosis. One daughter cell had gained something that the other daughter cell had lost.[...] These proved to be transposable elements [...]. It had to be discovered accidentally. »

Barbara McClintock, Nobel lecture, 8 December 1983

1- Barbara McClintock : la découverte des éléments mobiles

Barbara McClintock a reçu le prix Nobel de physiologie et de médecine en 1983 pour la découverte des éléments transposables (ET). Une expérience menée 40 ans auparavant, en 1944, donna lieu à l'observation de « phénotypes bizarres » non compatibles avec les principes de l'hérédité mendélienne, suscitant de fait le questionnement d'un changement de paradigme (Figure 1) (McClintock, 1944). L'été 1931 marqua le début de ses recherches en cytologie sur les cassures des extrémités chromosomiques chez le maïs. Barbara McClintock évoque la notion de gène comme un assemblage linéaire de petites perles (« beads ») le long d'un fil chromosomique. Néanmoins, cette époque correspond au développement de nouvelles méthodes qui permettaient d'étudier en détails l'état et la ségrégation des dix chromosomes du maïs. Par ce biais, elle mit en évidence pour la première fois et de manière précise, l'état des chromosomes au cours de la mitose et les différents types de réarrangements qui en découlent (translocations, inversions *etc.*).



Figure 1. Découverte des éléments mobiles par Barbara McClintock

Barbara McClintock fit la découverte de l'élément mobile *Ac* et l'élément non autonome *Ds* suite à l'étude des mosaïques des patrons de couleurs des semences de maïs. Son microscope et des épis de maïs présentant diverses mutations ont été exposés au National Museum of Natural History de Washington DC.

L'utilisation des rayons X par H. J. Muller et Hanson (Hanson and Heys, 1928; Muller, 1925; Muller, 1928) à la fin des années 20, chez la drosophile et parallèlement par Stadler chez l'orge, marqua profondément le monde des généticiens. L'irradiation des gamètes mâles (grain de pollen haploïdes) par les rayons X, permit à Stadler d'obtenir une grande collection de lignées mutantes, à laquelle était associée une large diversité de phénotypes (Stadler, 1928). Stadler proposa à Barbara McClintock de déterminer quels événements pouvaient causer l'émergence de phénotypes si particuliers. La première surprise fût qu'aucune des mutations

sélectionnées ne pouvait être associée aux changements phénotypiques observés. En revanche, chacune des anomalies coïncidait avec la perte d'un segment chromosomique. Ces résultats ont donné lieu à une observation minutieuse des réarrangements chromosomiques, notamment ceux liés à la perte de segments de chromosomes.

L'échange physique entre chromosomes homologues a ainsi été décrit pour la première fois, de même que l'organisation et le fonctionnement du centromère au cours d'une division cellulaire (McClintock, 1930). Suite aux cassures double brin causées par les radiations, Barbara McClintock identifia une structure particulière: les chromosomes en anneaux. Elle montra par la suite que cette structure dérivait de la fusion de deux chromosomes homologues au niveau de leurs extrémités cassées. Les radiations aux rayons X donnaient lieu à une grande diversité phénotypique au niveau des feuilles (McClintock, 1932). Elle démontra que ces phénomènes de variégations étaient systématiquement liés à la perte de ces anneaux dont la taille dépendait de la localisation du site de cassure. Aujourd'hui, on sait que les chromosomes en anneaux peuvent se former de manière spontanée dans certaines cellules (Figure 2) (Heim et al., 1987).



Figure 2. Caryotype d'une cellule cancéreuse issue d'un liposarcome

La présence de chromosomes en anneaux peut être utilisée comme outil diagnostique et permet, par exemple, de caractériser et classer les sous-types de sarcomes. Issu de (Heim et al., 1987).

Par exemple, dans une population de cellules sénescents, on observe un raccourcissement de la taille des télomères qui conduit à la formation de chromosomes en anneaux à plus forte fréquence (Gisselsson et al., 1999). Ces anomalies caryotypiques sont aussi observées au sein de populations de cellules cancéreuses et leur détection sert d'outil diagnostique. Au-delà de ces aspects appliqués, avant 1940, Barbara McClintock avait déjà contribué à l'émergence de concepts fondamentaux et quelques peu révolutionnaires sur l'organisation structurale du génome d'une cellule. Ce travail, réalisé au cours de treize années, avait généré la matière expérimentale propice à la «découverte accidentelle» des éléments de contrôle, les éléments transposables (ET). Toutes les souches issues d'un croisement unique présentaient une extraordinaire variété de patrons de couleurs (Figure 1), cela malgré leur fort apparemment génétique et malgré l'absence de traitement aux rayons X. Barbara McClintock conclut qu'un traumatisme génomique aurait provoqué l'activation d'éléments à l'origine des changements d'expressions de ces caractères phénotypiques (Emerson, 1914; McClintock, 1944). Elle démontra que le traumatisme initial était une rupture de l'extrémité d'un chromosome chez les deux parents. Ce qui contrastait avec les cassures aléatoires induites par rayons X, était la perte répétée du même fragment sur le chromosome 9, un site qu'elle nomma *Ds* (locus de dissociation). Elle identifia la présence d'un second élément

nécessaire à l'activation de cette cassure, le locus *Ac* (Activator). Elle nota par ailleurs que, suite à l'événement initial (lié à la présence du locus *Ac*), l'élément *Ds* se retrouvait à divers endroits sur le chromosome 9 et que ces mouvements d'excisions et de réinsertions corrélaient parfaitement avec les changements phénotypiques. En 1947, elle comprit que ces mutations «instables» étaient liées à la transposition non autonome d'un élément génétique (*Ds*) causée par la présence de l'élément autonome (*Ac*). Jusqu'alors les phénotypes mosaïques observés chez le maïs étaient attribués à l'intervention d'un facteur nucléaire qui inhibait transitoirement l'expression d'un gène de pigmentation (le gène *Dotted*) (Emerson, 1914; Rhoades, 1938). On sait, aujourd'hui que l'élément *Ac* est un transposon de 4,5 Kb qui code pour une transposase unique (Fedoroff, 1989). Il a aussi été montré que cette transposase permet la mobilisation de l'élément non autonome *Ds* (un dérivé de *Ac* possédant une délétion interne).

Cette découverte a longtemps été perçue comme marginale et il est étonnant de constater que ses travaux ont mis près de 40 années avant d'être admis par la communauté scientifique. D'autant plus étonnant qu'avant cette découverte, McClintock était déjà hautement considérée par ses pairs (Botanical Society of America Merit Award 1957, National Academy of Science Kimber award 1967). La reconnaissance de ses travaux est finalement passée par la redécouverte des éléments transposables dans les années 70, d'abord chez les bactéries puis dans les organismes modèles des généticiens (Bucheton, 1973; Campbell, 1962; Kidwell et al., 1977; Shapiro, 1979). Très tôt après la publication de l'article qui date la découverte des transposons (McClintock, 1950), les généticiens du maïs ont compris l'impact de ces résultats, comme peuvent en témoigner les publications d'un certain nombre de chercheurs dans le domaine (Brink and Nilan, 1952; Dollinger, 1954; Kreizinger, 1960). Il semble néanmoins, que ces répercussions soient restées cloisonnées à ce petit monde de spécialistes. Lorsqu'en 1951, McClintock donne un séminaire au Cold Spring Harbor Symposium, la communauté scientifique reste de marbre devant la quantité et la complexité des résultats présentés. Evelyn Witkin témoigna plus tard «there was a dead silence -- a foretaste of the initial reception her finding would receive» (McGrayne SB, 2001, in *Nobel Prize Women in Science*). Ce silence fait aussi écho au paradoxe entre le concept d'information génétique contenu par des «éléments de contrôle» mobiles et l'idée que l'information génétique est contenue par une molécule stable, l'ADN. En effet, en 1944, Oswald Avery montre chez les pneumocoques, que la digestion de l'ADN empêche la transformation bactérienne et qu'il s'agirait donc du support de l'hérédité. En 1952, Hershey et Chase montrent à leur tour que le matériel transmis et capable d'infecter *Escherichia coli*, est bien l'ADN du phage T2 et non ses protéines. Enfin, Watson, Crick et Franklin, obtiennent en 1953 un cristal d'ADN en double hélice par diffraction aux rayons X. Les répercussions de ces découvertes sur la communauté scientifique ont vraisemblablement créé un contexte défavorable à la compréhension des éléments transposables. A travers diverses techniques et organismes modèles, la communauté scientifique s'était forgée un faisceau solide de preuves qui validait définitivement le paradigme de l'hérédité mendélienne dont l'objet central était l'ADN en tant qu'entité stable. Comment, dans ce contexte, le support stable de l'information génétique pouvait-il être la source même de sa propre instabilité ? Ceci suscite

par ailleurs un questionnement quant à l'impact d'un modèle dominant en science, notamment lorsque l'observation contredit les attentes «prédites» par le modèle.

McClintock avait fini par déduire certaines propriétés du génome et de la cellule, sans véritable preuve expérimentale cependant. C'est ce qu'elle exprime dans son discours de prix Nobel «*The significance of responses of the genome to challenge*». L'idée générale est celle d'un génome extrêmement dynamique, dont l'instabilité nécessitait l'intervention de mécanismes cellulaires «post-traumatiques» extrêmement fins. Cette vision déterministe (voire finaliste) d'une cellule intelligente («*smart cell*») est le fruit de plusieurs décennies d'expérimentations passées à retracer minutieusement la rupture des chromosomes au cours du développement pour des dizaines de milliers de plants de maïs. Des décennies au cours desquelles, McClintock restait fascinée par la capacité de résilience des organismes, suite à un véritable traumatisme génomique. McClintock restait aussi pleinement consciente des myriades de mécanismes inconnus qui ne pourront être révélés que par des observations encore insoupçonnées.

2- Les plasmides et les phages : les épisomes

A la fin des années 40, la communauté des microbiologistes est intriguée par les mécanismes d'échanges génétiques chez les bactéries. Ces études ont conduit à la description de la conjugaison bactérienne et de la transduction de l'ADN *via* les bactériophages, deux mécanismes au cours desquels des séquences d'ADN mobiles, «les épisomes», ont la capacité à s'intégrer à différentes positions dans le génome hôte.

Les travaux sur la conjugaison bactérienne ont été initiés par la découverte de la recombinaison chez *E.coli* en 1946 par Lederberg et Tatum (Lederberg and Tatum, 1946). En 1952, William Hayes a montré qu'en absence de recombinaison, on observait tout de même un transfert de matériel génétique unidirectionnel, d'une souche donneuse à une souche receveuse. Il conclut qu'un facteur infectieux (*F factor*) avait été transmis et que ce dernier devait se répliquer indépendamment et plus rapidement que le chromosome bactérien: la notion de plasmide (*F plasmid*) venait d'être définie pour la première fois (Hayes, 1952). Certaines souches donneuses ont pu être isolées pour leur capacité à transférer plus de matériel que les souches standards, ces souches découvertes en 1961, furent nommées *Hfr* (*High frequency of recombination*). L'étude des souches *Hfr* par Elie Wollman et François Jacob, a montré que ces transferts de matériel génétique résultaient de l'intégration du plasmide *F* au sein du génome bactérien. Ce dernier possédait la capacité de se répliquer de manière autonome, tout en répliquant une partie du chromosome du génome hôte (Figure 3).

Parallèlement, Zinder et Lederberg (Zinder and Lederberg, 1952) d'une part et Morse (1954) d'autre part, ont montré chez *E. coli* et *Salmonella*, que certaines particules de phages pouvaient transférer des fragments du chromosome bactérien d'une cellule à une autre. Ce processus de transduction décrit en fait l'encapsulation d'ADN bactérien qui s'intègre au génome d'une autre bactérie au cours des cycles d'infection

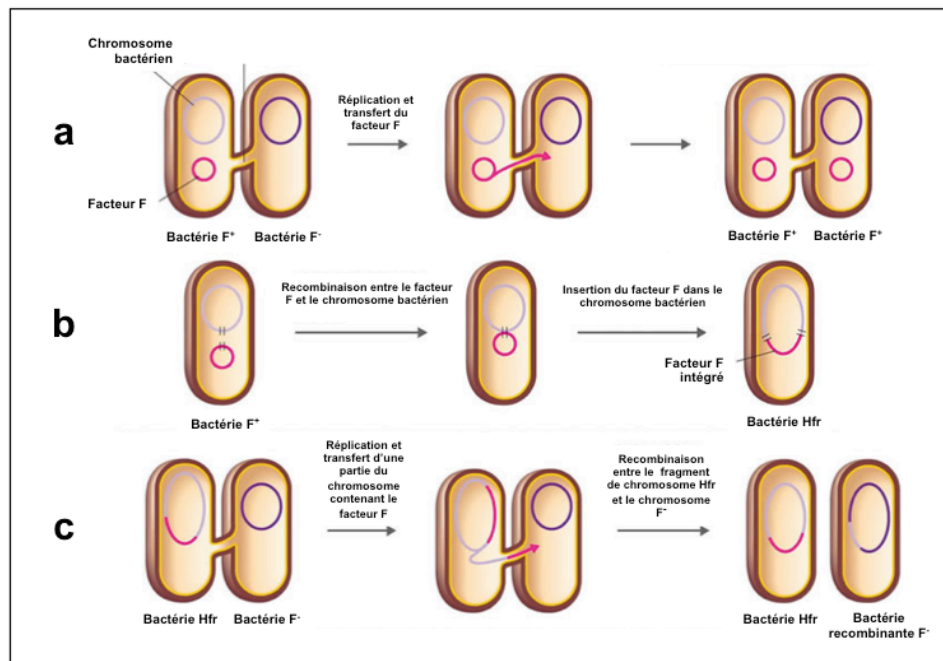


Figure 3. Le transfert du facteur F chez les bactéries

(a) Le facteur F (plasmide) d'une bactérie F^+ est transféré à une bactérie receveuse F^- qui est convertie en une bactérie F^+ .
 (b) Dans certains cas, le plasmide F peut s'intégrer dans le chromosome de la bactérie. La souche est alors appelée Hfr pour *High frequency of recombination*.
 (c) Une bactérie Hfr peut transférer à une bactérie F^- une partie de son chromosome contenant un fragment de facteur F intégré. Ceci génère une nouvelle souche recombinante F^+ . © Copyright 2010 Pearson Education, Inc.

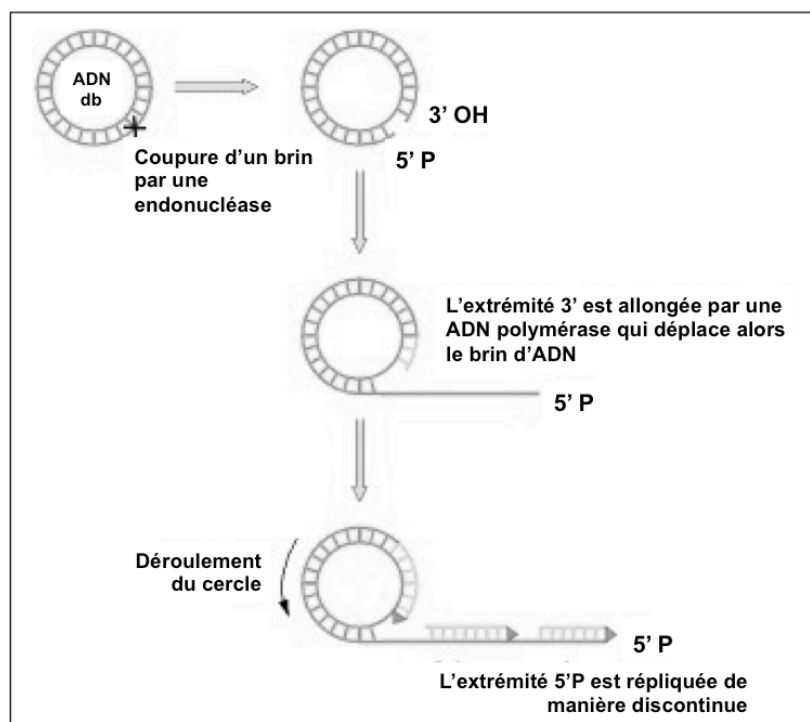


Figure 4. Mécanisme de réplication en cercle roulant

L'ADN circulaire double brin est coupé par une endonucléase sur un de ses brins libérant une extrémité 3' OH et une extrémité 5' P. L'extrémité 3' OH est alors répliquée par la polymérase III par complémentarité sur le brin matrice non coupé. La réplication à l'extrémité 3' OH déplace alors le brin d'ADN à l'extrémité 5' P créant ainsi un mécanisme de cercle roulant. L'extrémité 5' P est également répliquée de manière discontinue par la polymérase. La réplication peut se faire de manière continue produisant de multiples copies à la suite (appelé concatémère). Ces copies peuvent ensuite être converties en ADN double brin circulaire. Adapté de Genetics: Principles and Analysis, 4th ed. Jones and Bartlett, 1998.

du phage. En 1962, Campbell fait l'hypothèse du mode de réplication en *rolling circle*, qui impliquerait la reconnaissance de sites spécifiques complémentaires pour sa recombinaison et son excision du chromosome hôte. Il expose ainsi les prémices des mécanismes pour la reconnaissance, l'excision et l'intégration d'une molécule d'ADN exogène dans le génome. On sait aujourd'hui que certains transposons comme les *Helitrons*, transposent *via* ce mécanisme de réplication en cercle roulant (*rolling circle*) (Figure 4).

3- La convergence des arguments : la notion étendue d'éléments génétiques mobiles

Jusque dans les années 60, le rapprochement conceptuel entre les éléments de contrôle du maïs et les épisomes bactériens n'était pas perçu par la communauté. Les exemples d'éléments génétiques mobiles étaient perçus comme des exceptions biologiques issues des réarrangements aberrants et d'événements de «recombinaisons illégitimes». Le développement de la biologie moléculaire par l'utilisation des phages et des plasmides *in vivo* renforça cependant l'idée que le génome pouvait intégrer des éléments exogènes sans que cela soit nécessairement délétère pour l'hôte. La découverte d'éléments mobiles ayant pour seule fonction leur capacité à se mobiliser d'un endroit à l'autre du génome marqua les esprits (Taylor, 1963). La conférence d'Asilomar en 1975, pendant laquelle sont présentés ces travaux (Shapiro, 1979; Taylor, 1963), symbolise l'acceptation de l'idée que les transposons font partie intégrante du génome. Le bactériophage *Mu*, découvert en 1963 par Taylor (Taylor, 1963), a été nommé ainsi car il causait des mutations à un grand nombre de sites sur le chromosome d' *E. coli*. Par ailleurs, ces séquences semblaient pouvoir s'insérer à n'importe quel locus dans le génome. Plus tard, Shapiro démontra le mécanisme de transposition par lequel *Mu* s'insère dans le génome (Figure 5) (Shapiro, 1979). En 1967, Shapiro s'intéresse à la régulation de l'opéron lactose. Il isole un certain nombre de mutations spontanées qui ne ressemblaient en aucun point à des délétions, aux mutations par substitution ou par décalage du cadre de lecture. Il conclut que l'insertion d'un fragment d'ADN était responsable de ces mutations: c'est la découverte des séquences *IS* (Insertion Sequences), à savoir, différents fragments d'ADN insérés à différents sites dans l'opéron (Shapiro, 1969). L'importance biologique des séquences *IS* prit une autre dimension lorsqu'elles furent identifiées dans les plasmides (e.g. le plasmide *F*) et associées aux sites de recombinaisons entre le plasmide et le chromosome bactérien.

La recherche des facteurs génétiques impliqués dans la résistance aux antibiotiques, vectorisée par les plasmides, s'intensifie au début des années 60. Nombre de ces déterminants présentaient des profils de recombinaison anormaux. Ces déterminants avaient la capacité de transposer d'un réplicon à l'autre et étaient responsables de la résistance induite par transfert de matériel génétique. Dans les années 70, il a été montré que des séquences codantes pouvaient être vectorisées par les extrémités inversées répétées des *IS* (e.g. *Tn3* pour la résistance à l'ampicilline). Il est alors admis que les transposons peuvent conférer un avantage sélectif à l'hôte ce qui pourrait en expliquer leur maintien dans les génomes (Hedges and Jacob, 1974; Heffron et al., 1975).

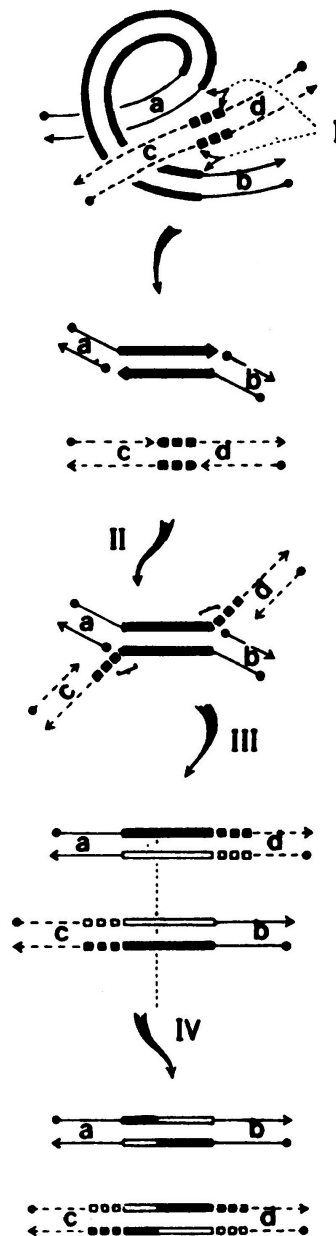


Figure 5. Modèle de transposition du bactériophage Mu

Cette légende est la traduction de la légende originale publiée en 1979 (Shapiro JA, 1979, PNAS). « Le schéma du haut illustre comment plusieurs régions d'ADN peuvent être rapprochées physiquement suite à des événements de clivages et de ligations. Les quatre schémas du bas montrent les différentes étapes dans le processus de transposition. Les lignes pleines indiquent les brins d'ADN donneurs, et les lignes pointillées l'ADN receveur. Les rectangles noirs correspondent à l'ADN parental duquel l'élément transposable est issu, et les rectangles gris correspondent à l'ADN nouvellement synthétisé. Les lettres a, b, c, d, dans les bras flanquants l'élément transposable et l'oligonucléotide cible servent à indiquer la structure génétique des différents duplex produits. »

Il s'agit de l'un des premiers modèles pour expliquer un mécanisme de transposition, et notamment l'un des plus parcimonieux. Ici, les simples brins d'ADN donneurs sont directement ligués à leur cible de manière à créer la jonction entre les deux molécules. L'intermédiaire ainsi généré crée deux fourches de réplifications à chaque jonction (expliquant comment *Mu* transpose en même temps qu'il est répliqué). Le modèle permet également d'expliquer de manière simple les inversions et les délétions au niveau des sites d'insertion de l'élément. Ce modèle est le premier à suggérer l'existence des ET en tant qu'entité autonome.

Il apparaît peu à peu, au sein de la communauté, que la mobilité de séquences à ADN est un caractère commun à tous les génomes, également présent chez les organismes modèles «complexes». La découverte des éléments *P*, *hobo* et *FB* chez la drosophile, de même que l'élément *Tc1* chez *Caenorhabditis elegans* suggérait que ces éléments se mobilisaient par des mécanismes similaires. L'activité biochimique responsable de leur mobilité passait par la reconnaissance d'une structure particulière de l'ADN, suivie d'une coupure et d'une réinsertion à un autre site du génome. Les spécificités mécanistiques et leur diversité renseignaient quant à leurs origines indépendantes. Néanmoins, tous étaient capables de se couper puis de se coller au sein d'une molécule d'ADN. C'est ainsi que ce type d'élément fut associé à la même grande classe, les transposons de Classe II (Figure 6 et 6bis).

Il s'est avéré plus tard que la Classe d'éléments mobiles la plus représentée chez la levure, la drosophile et les vertébrés, transposait *via* un intermédiaire ARN rétro-transcrit, avant d'être insérée dans le génome : ce sont les éléments de Classe I. Structurellement et biochimiquement, on distingue deux types d'éléments transposables de Classe I (Figure 6). Le premier type comprend les éléments apparentés aux rétrovirus (*e.g.* les *Ty* chez la levure), auxquels on peut attribuer la majorité des mutations spontanées visibles chez la drosophile (*e.g.* *copia*, *412*, *B104*, et *gypsy*). Ces éléments ont été associés à l'émergence de certaines pathologies, comme par exemple, lors de la découverte d'insertions spécifiques au niveau des gènes proliférations cellulaires (oncogènes). Par la suite, un autre groupe majoritaire a été caractérisé, il s'agit des *LINE* (Long Interspersed Nucleotide Elements) et des *SINE* (Short Interspersed Nucleotide Elements) qui ne possèdent pas de LTR (Long Terminal Repeats), contrairement aux autres rétrotransposons (Figure 6).

4- Les éléments transposables, des facteurs majeurs de l'évolution

Avant la découverte des éléments transposables (ET), les génomes eucaryotes étaient considérés comme stables, peu variables en termes de taille et de contenu. Le fait que les ET puissent constituer une partie non négligeable du génome, potentiellement impliquée dans une grande diversité de remaniements génétiques permit de reconsidérer leur rôle délétère au cours de la vie d'un organisme et au cours de l'évolution. Outre leur insertion dans les gènes, plusieurs exemples d'inversions, de recombinaisons ectopiques ont été associés à la mobilisation des éléments transposables. Parmi ces exemples, l'étude de l'élément *P* chez *D. melanogaster* retrace historiquement, la prise de conscience de ces nouvelles considérations.

La mobilisation de l'élément *P* a été associée à un développement aberrant des gonades, accompagnée d'une stérilité des individus. Les premières observations de ce type de phénotypes datent des années 70, sans que cela soit directement associé à l'élément *P* (Hiraizumi, 1971). Leur expérience consistait en une série de croisements entre des souches fraîchement prélevées dans la nature et d'anciennes souches de laboratoire. La descendance de ces croisements avait notamment subi des événements de recombinaisons chez les mâles, un phénomène jamais observé jusqu'alors chez *D. melanogaster*. Cette observation fût la première parmi une série

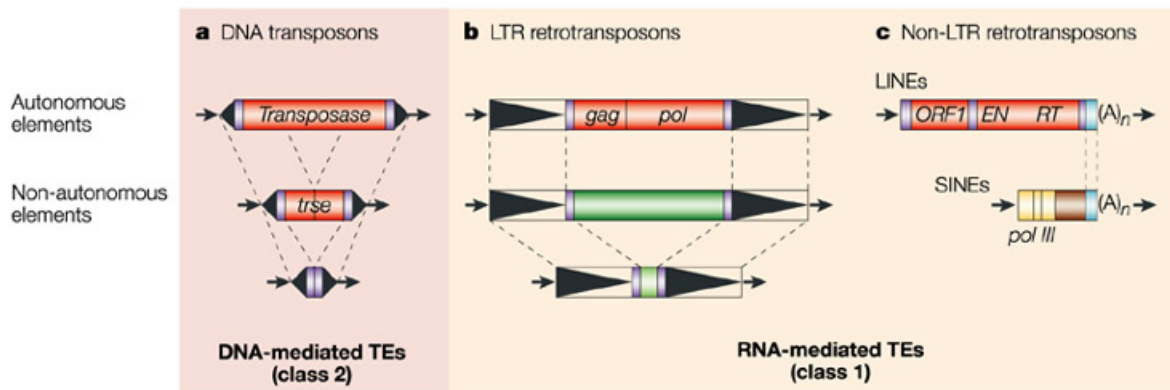


Figure 6. Les deux Classes d'éléments mobiles

La classification divise les ET en deux Classes principales sur la base de la présence ou l'absence d'un intermédiaire ARN pour la transposition. La Classe I contient les éléments transposable sans intermédiaire ARN (transposons à ADN) (a). La Classe des rétrotransposons contient majoritairement les éléments transposables de l'ordre des LTR (b), et celui des *LINE* et des *SINE* qui ne contiennent pas de LTR (c). Issu de (Feschotte et al., 2002).

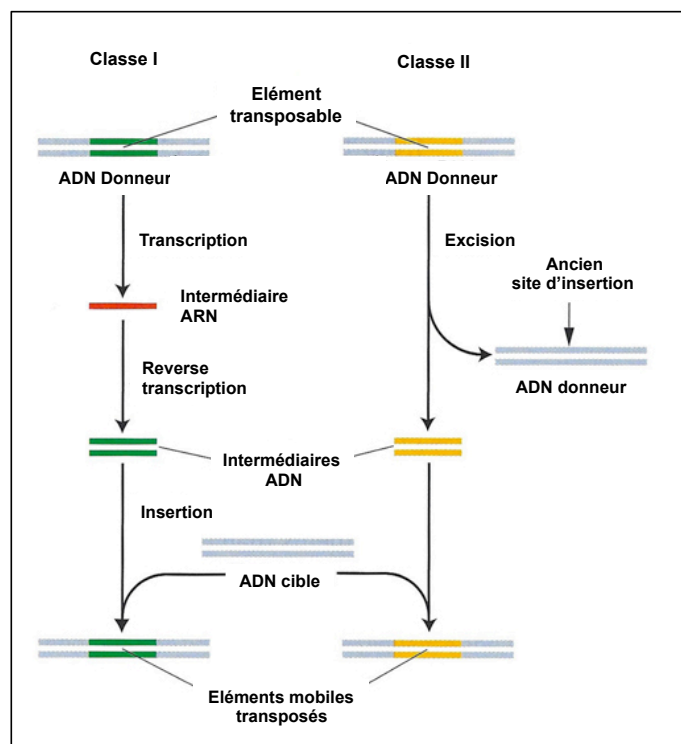


Figure 6bis. Les mécanismes de transposition des deux Classes d'éléments mobiles

Les éléments mobiles sont divisés en deux Classes. Les éléments mobiles de Classe I sont copiés en deux étapes. Dans un premier temps, ils sont transcrits d'ADN en ARN. L'ARN produit subit ensuite une *reverse transcription* pour générer un intermédiaire ADN qui est alors inséré dans la séquence d'ADN cible. Cette *reverse transcription* est catalysée par une *reverse transcriptase* qui est très souvent codée par l'élément mobile lui-même. Les éléments de Classe II transposent *via* un mécanisme de type « couper-coller » mais n'implique pas d'intermédiaire ARN. L'élément mobile est excisé de l'ADN donneur et inséré dans une séquence d'ADN cible. Les mécanismes de réparation de l'ADN peuvent dupliquer tout ou partie de l'élément depuis son site d'excision. Adapté de (Paul, 2013) Biol202 (photograph).

d'autres qui paraissaient déconnectées, mais toutes issues du même type de croisement dont la descendance présentait chaque fois des aberrations phénotypiques telles qu'une distorsion de ségrégation, une stérilité ou un taux élevé de mutations. En 1977, Kidwell attribua l'ensemble de ces anomalies à un syndrome qu'elle nomma «dysgenèse hybride» (Kidwell et al., 1977) (Figure 7). Kidwell classa les souches en deux catégories, M et P (Maternelle et Paternelle) dont la descendance stérile était systématiquement issue du croisement d'une femelle M par un mâle P (et non du croisement réciproque). Et quand bien même ces hybrides étaient fertiles, ceux-ci présentaient le type d'anomalies phénotypiques précédemment décrites. La conclusion fût que les lignées germinales des hybrides M x P étaient instables. En 1979, Engels montra que le phénotype dysgénique était associé à des «facteurs» présents à diverses positions dans le génome de la souche P. Ces «facteurs *P*» agissaient de manière indépendante et leur distribution dans le génome variait d'une souche à l'autre. En 1981, Engels montra que 85% des points de cassures chromosomiques induits au cours de la dysgenèse étaient limités à un petit nombre de loci (Engels, 1981; Engels and Preston, 1981). Ces loci n'étaient jamais présents sur les chromosomes issus de la souche M, suggérant que ces points correspondaient aux sites d'insertions du «facteur *P*». Cette observation a apporté une indication importante quant à la nature de ce dernier, vraisemblablement une famille d'élément transposable présente dans les souches P mais absente chez les souches M. La mobilisation de ces éléments dans les hybrides semblait être responsable des anomalies phénotypiques du syndrome de dysgenèse hybride. L'hypothèse du facteur *P* a par la suite été confirmée par l'identification d'une insertion dans le locus *white* (Rubin et al., 1982). Il a été montré que cette insertion correspondait bien à un site de cassure chromosomique et qu'il s'agissait de l'élément transposable *P*, ainsi caractérisé pour la première fois au niveau moléculaire (Bingham et al., 1982; Rubin et al., 1982). Hormis l'exemple de *P*, pour lequel les sites d'insertions étaient décrits avec une résolution fine, la question de la spécificité des sites d'insertions dans le génome restait peu accessible. Néanmoins, dans les années 80, la caractérisation de nouveaux rétrotransposons (tels que *17.6*, *297*, *Beagle*, *Gypsy*, *Copia* etc.) a montré leur prédisposition à s'insérer dans les régions riches en AT, notamment dans les séquences promotrices des gènes (Saigo et al., 1984; Snyder et al., 1982).

D'autres systèmes associés à un syndrome de dysgenèse hybride ont été décrit en parallèle des études menées sur *P*. Par exemple, le système I-R, lié à la présence de l'élément *I*, a été caractérisé un peu après la découverte du facteur *P*, bien qu'au début des années 70 un facteur de stérilité femelle non mendélien avait été décrit (Bucheton, 1973). Plus tard, la description du système *hobo* fait aussi référence dans le cadre de la dysgenèse hybride (Streck et al., 1986; Yannopoulos et al., 1987). L'une des conclusions apportée par ces travaux était que les éléments *P*, *I* et *hobo* avaient été introduits ou réintroduits chez *D. melanogaster* par transferts horizontaux ((Daniels et al., 1990) pour l'élément *P*). Ces trois exemples, sont associés à l'idée que les ET peuvent causer d'importants remaniements, souvent hautement délétères pour l'hôte, lorsqu'ils émergent *de novo* dans une espèce. Cela a notamment donné lieu aux théories développées par Charlesworth qui supposent que les ET sont régulièrement purgés du génome par la sélection naturelle (Charlesworth and Langley, 1986).

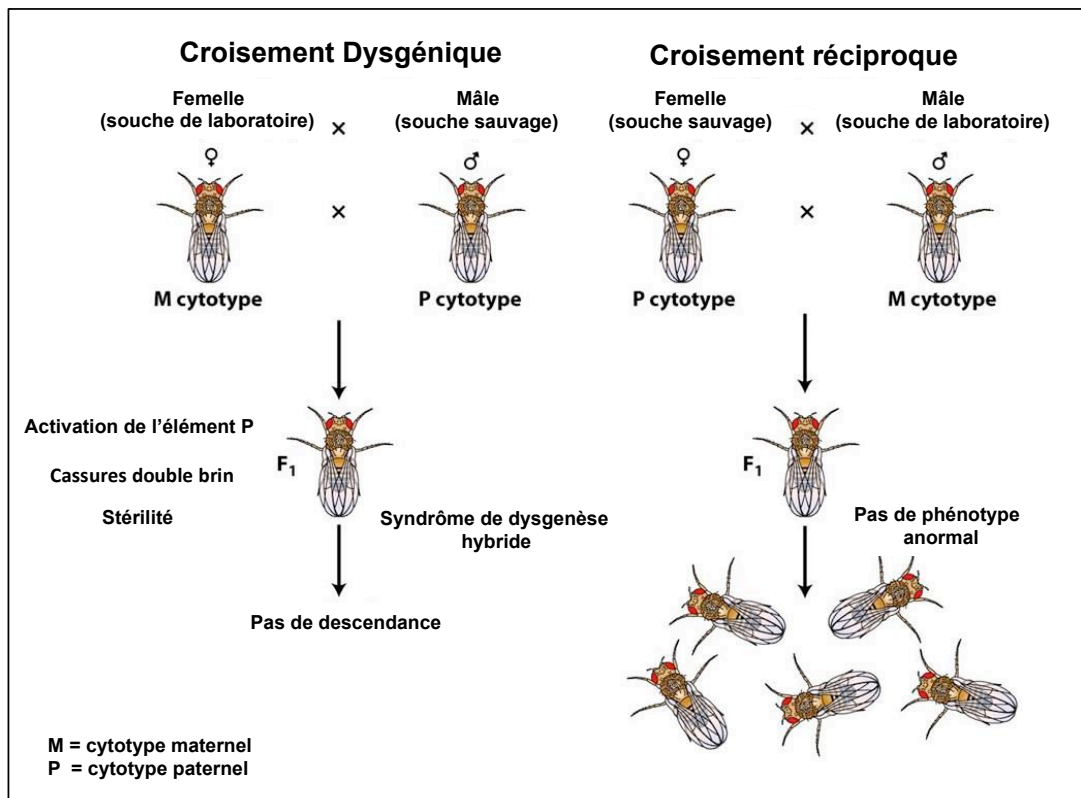


Figure 7. Dysgénèse hybride

La dysgénèse hybride a été découverte la première fois en laboratoire, en observant la descendance du croisement entre des femelles de laboratoire de cytotype M et des mâles sauvages de cytotype P. Les mouches issues de ce croisement présentaient des forts taux de cassures double brin, une activation de l'élément P et une stérilité de leur descendance. Au contraire, dans un croisement réciproque impliquant des femelles sauvages de cytotype P et des mâles de laboratoire de cytotype M, aucun phénotype anormal n'est observé dans la descendance. Adapté de Introduction to Genetic Analysis, Ninth Edition © W.H Freeman and Company.

5- La contribution des modèles : «du neutralisme au sélectionnisme»

Devant les limitations expérimentales, les théoriciens de la génétique quantitative et de la génétique des populations ont modélisé différents scénarios pouvant expliquer l'invasion et le maintien des éléments transposables dans le génome. Depuis la synthèse moderne de l'évolution (l'idée d'une évolution graduelle façonnée par la sélection naturelle) développée par un consortium de chercheurs dans les années 40 (e.g. Dobzhansky, T. 1937. *Genetics and the origin of Species* ; Huxley, 1942. *Evolution : The modern synthesis*), une avancée conceptuelle avait été apportée par Motoo Kimura. C'est la théorie neutraliste de l'évolution moléculaire. Kimura suggère que la majorité des différences moléculaires entre espèces ou populations est sélectivement neutre et qu'il est difficile de l'expliquer par l'effet de la sélection naturelle (Kimura, 1979). Il suggère que ces différences sont essentiellement expliquées par une distribution aléatoire d'allèles soumis à la dérive génétique. Cette affirmation donna lieu à une véritable controverse au sein de la communauté scientifique.

Les premiers modèles théoriques (Kimura, 1981; Ohta and Kimura, 1981) étaient basés sur une vision neutraliste de l'établissement des éléments transposables dans un génome. Ce modèle repose sur l'observation que, lorsque le nombre de copies augmente en moyenne dans une espèce, les différences entre populations augmentent aussi considérablement. La dérive et les variations de la taille efficace des populations étaient les paramètres principaux pour décrire le gain ou la perte des ET dans les populations. Les gains ou pertes de fitness associés aux mutations, ainsi que les événements de migration, la fréquence des transferts horizontaux et le rapport duplications/délétions des ET, étaient aussi des paramètres implémentés ou discutés autour de cette hypothèse neutraliste.

En opposition à cette vision neutre de l'évolution, il a été proposé que le maintien des ET pouvait s'expliquer par l'avantage direct qu'il conférerait à leur hôte. Cette hypothèse est largement issue des exemples bactériens montrant que les ET avaient la capacité de transférer horizontalement de «bons gènes» (Hedges and Jacob, 1974; Heffron et al., 1975). Le gain de fitness apporté par les transposons vecteurs de résistances aux antibiotiques, soutenait vivement l'idée que leur maintien était indispensable à la survie de l'hôte, en particulier dans des conditions de stress extrêmes. Dans la même ligne d'idées, les transposons étaient perçus comme des vecteurs de variabilité génétique, une propriété essentielle à l'hôte pour son adaptation aux changements environnementaux et son évolution.

Indépendamment de ses deux visions, le modèle de «transposition répllicative» proposé par Hickey (Hickey, 1982), ne fait aucune hypothèse quant à l'effet de la sélection naturelle sur la présence des ET. Il propose simplement que les transposons se multiplient plus rapidement que les autres composants du génome et que cela suffit à expliquer leur maintien, c'est l'idée que les ET se comportent comme des «gènes égoïstes». Néanmoins, le modèle de transposition répllicative dans un contexte évolutif neutre, prédit une augmentation exponentielle du nombre de transposons par rapport aux autres éléments du génome. Cela doit conduire, à terme, à une explosion du nombre de copies pour chacune des familles présentes dans le génome, ce qui est en contradiction avec l'observation (e.g. 30 à 50 copies de *P* chez *D. melanogaster*). Ce modèle implique donc qu'au delà d'un certain seuil, la sélection opère pour éliminer des copies et/ou pour en limiter l'expansion. Ces contraintes évolutives sont associées aux effets délétères induits d'une part *via* de nouvelles insertion dans les gènes et d'autre part *via* des événements de recombinaisons ectopiques, dont le taux augmente quadratiquement avec le nombre de copies (Doolittle and Sapienza, 1980; Orgel and Crick, 1980).

6- Les éléments égoïstes auto-régulés

Pour étudier la dynamique des ET selon le modèle neutre, Deborah et Brian Charlesworth proposent, en 1983, le concept d'auto-régulation (Charlesworth, 1983). C'est-à-dire que le taux de transposition par copie et par génération diminue à mesure que le nombre de copies augmente dans le génome. Dans ce cadre, lorsque le taux de transposition est élevé et le nombre de copies faibles, l'élément peut envahir le génome jusqu'à ce que le

taux de transposition soit égal au taux d'élimination de l'élément, point auquel le nombre de copies atteint un état d'équilibre dans le génome. Cette hypothèse prédit que les ET maintenus dans le génome sont ceux sélectionnés pour leur capacité à limiter leur expansion. Malgré le manque de données expérimentales, ce modèle a été validé en théorie dans les systèmes procaryotes et sous certaines conditions pour les espèces sexuées (Charlesworth and Langley, 1986). Pour une espèce sans contrainte pour la recombinaison et avec un nombre élevé de copies, il est prédit que les chances de survie sont meilleures lorsque les ET présents possèdent une capacité intrinsèque d'auto-régulation. D'après ce modèle, un tel type d'ET est sélectionnable à condition que les effets délétères induits soient suffisamment élevés. Historiquement, ce modèle était adapté aux effets extrêmes, observés au cours du syndrome de dysgenèse hybride (taux de transposition très élevé, forte létalité ou/et stérilité chez la descendance). Malgré ces limitations, il s'agit du socle théorique à partir duquel la plupart des modèles suivants ont été développés. De plus, il est le premier à poser la problématique de la régulation comme limite à l'expansion des ET.

7- Base expérimentale et modèles pour la régulation

A - Inhibition somatique

La régulation des ET, considérée comme une stratégie évolutive stable, a été activement testée dans les années 90. Le premier exemple a été apporté *via* l'élément *P* dont l'expression est réprimée en lignée somatique (Chain et al., 1991; Laski et al., 1986; Laski and Rubin, 1989; Siebel and Rio, 1990). Suite à l'activation de l'élément *P* par un promoteur heat-shock, une isoforme longue a pu être isolée en lignée somatique uniquement. Ce transcrit long était caractérisé par la rétention d'un intron entre les exons 2 et 3 de la transposase et lié à la présence d'un facteur protéique inhibiteur empêchant l'épissage du pré-ARN messager. Du point de vue évolutif, les ET ne présentent pas d'avantage particulier à transposer en lignée somatique puisque qu'aucune des insertions ne sera transmise à la génération suivante. Au contraire, cela pourrait augmenter la fréquence des mutations délétères dans la population. En somme, l'inhibition somatique des ET peut constituer un *trade-off* limitant les effets délétères sur l'organisme tout en maintenant leur capacité à envahir la lignée germinale.

B - L'empoisonnement par les copies non-autonomes

Parallèlement d'autres modèles ont été proposés, essentiellement basés sur l'étude de transposons à ADN (les éléments de Classe II). Les trois modèles présentés ici, reposent sur l'hypothèse qu'en transposant, des copies non canoniques ou bien délétées peuvent être générées et ainsi entrer en compétition avec les copies actives. Le premier mécanisme correspond à une inhibition compétitive entre les formes canoniques et non canoniques pour l'accès aux ITR, qui par ce biais limiterait le taux d'excision de l'élément. La régulation par des isoformes polypeptidiques pour l'élément *P* (Lee et al., 1996; Rio and Rubin, 1988) ou des copies délétées

pour l'élément *mariner* (*Dominant Negative Complementation* (Lohe and Hartl, 1996)), convergent vers ce modèle. Le second modèle est l'hypothèse de titration, c'est-à-dire que les éléments non-autonomes auraient la capacité à utiliser la transposase produite par les éléments autonomes et par conséquent limiteraient la quantité de transposase disponible pour ces derniers. Le dernier mécanisme proposé pour les éléments *P*, *mariner* et *Ac* est l'idée du multimère poison, proposé après la démonstration que la transposase de ces éléments était active sous forme oligomérique (Lohe et al., 1997). L'association entre des monomères fonctionnels et d'autres non fonctionnels aurait pour conséquence la formation d'un multimère inefficace pour la transposition. Aucune de ces hypothèses n'a malheureusement pu être formellement démontrée ni génétiquement, ni biochimiquement (Townsend and Hartl, 2000).

C - Les prémices de la régulation par ARN interférences

Il faudra attendre la fin des années 90 et la découverte de l'ARN interférence pour révolutionner les modèles de régulation des ET. Les prémices de cette révolution avaient néanmoins été pressenties dès 1993. Raymond et Simmons montrent qu'un ARN orienté en antisens permettait la répression post-traductionnelle de l'élément *P*. Leur interprétation était que cet ARN antisens, produit par un promoteur externe, pouvait se lier à l'ARN sens de l'élément par complémentarité et bloquer le processus de traduction. L'élément *P* ainsi réprimé, était associé à la disparition de la létalité pupale et du syndrome de dysgenèse hybride (Simmons et al., 1996). L'autre fait important concernant la répression de la dysgenèse a été apporté par Ronsseray, Lehman et Anxolabéhère dès 1991 (Ronsseray et al., 1991; Ronsseray et al., 1996). Ils ont montré par des techniques de cytologie, qu'une insertion spécifique de l'élément *P* au niveau d'un télomère du chromosome X (position 1A) suffisait à réprimer complètement la dysgenèse hybride. Cette insertion située à proximité des régions télomériques, a été décrite comme un site préférentiel d'insertion pour l'élément. Par ailleurs, la transmission de cette insertion par la mère était suffisante pour annuler les effets délétères associés au syndrome de dysgenèse hybride (Ronsseray et al., 2001; Ronsseray et al., 1993; Stuart et al., 2002). Il s'est avéré plus tard, que cette insertion avait la capacité à réprimer l'élément *P* par des mécanismes d'ARN interférences. Ce type de preuves expérimentales a été facilité notamment par le développement de nouvelles méthodes de séquençage haut-débit vers la fin des années 90.

En 1995, le premier génome complet d'un organisme vivant (*Haemophilus influenzae*) a été complètement séquencé. Depuis, le nombre de génomes séquencés croît de manière exponentielle (Figure 8). En parallèle, un constat évident est fait quant à la nature des génomes: tous possèdent une fraction de séquences répétées au sein de laquelle on retrouve une large diversité d'éléments transposables. La production intensive de nouvelles données (Figure 9) à l'échelle du génome a conduit à reconsidérer la question du maintien des ET de manière plus objective. Ainsi, l'étude comparative de génomes qui possèdent des propriétés bien distinctes, a permis d'étayer un certain nombre d'hypothèses concernant les contraintes évolutives associées à la mobilisation des ET. Un aperçu est donné dans la section suivante.

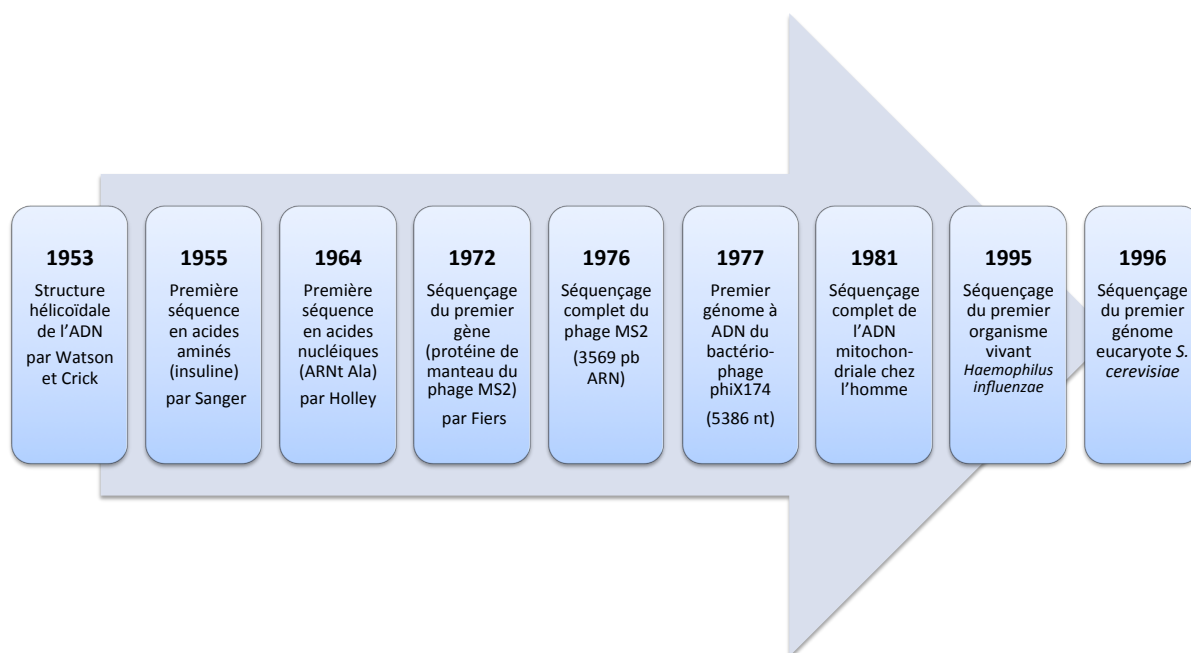


Figure 8. Chronologie des grandes avancées des techniques de séquençage

La publication de la structure hélicoïdale de l'ADN en 1953, suivie de la publication de la séquence en acides aminés de l'insuline en 1955 par Sanger, a canalisé les efforts pour séquencer les acides nucléiques, porteurs de l'information génétique. En 1964, la première séquence d'acide nucléique venait d'être obtenue par Robert W. Holley. Il s'agissait de l'ARNt alanine. Il faut attendre 1972 pour le séquençage du premier gène codant une protéine de manteau du phage MS2 par Walter Fiers, suivie du séquençage complet de ce phage (3569 paires de bases ARN) en 1976. Le développement des techniques de séquençage par Sanger (et Gilbert) permis en 1977 d'obtenir le premier génome à ADN, les 5386 nucléotides simple brin du bactériophage phiX174. Cette période correspond à l'intensification des projets de séquençages de génomes, qui donne lieu en 1981 au génome complet de la mitochondrie chez l'homme, puis au premier génome de chloroplaste en 1986. En 1995, c'est le séquençage du premier organisme vivant, *Haemophilus influenzae*, puis du premier génome eucaryote, *S. cerevisiae*, en 1996.

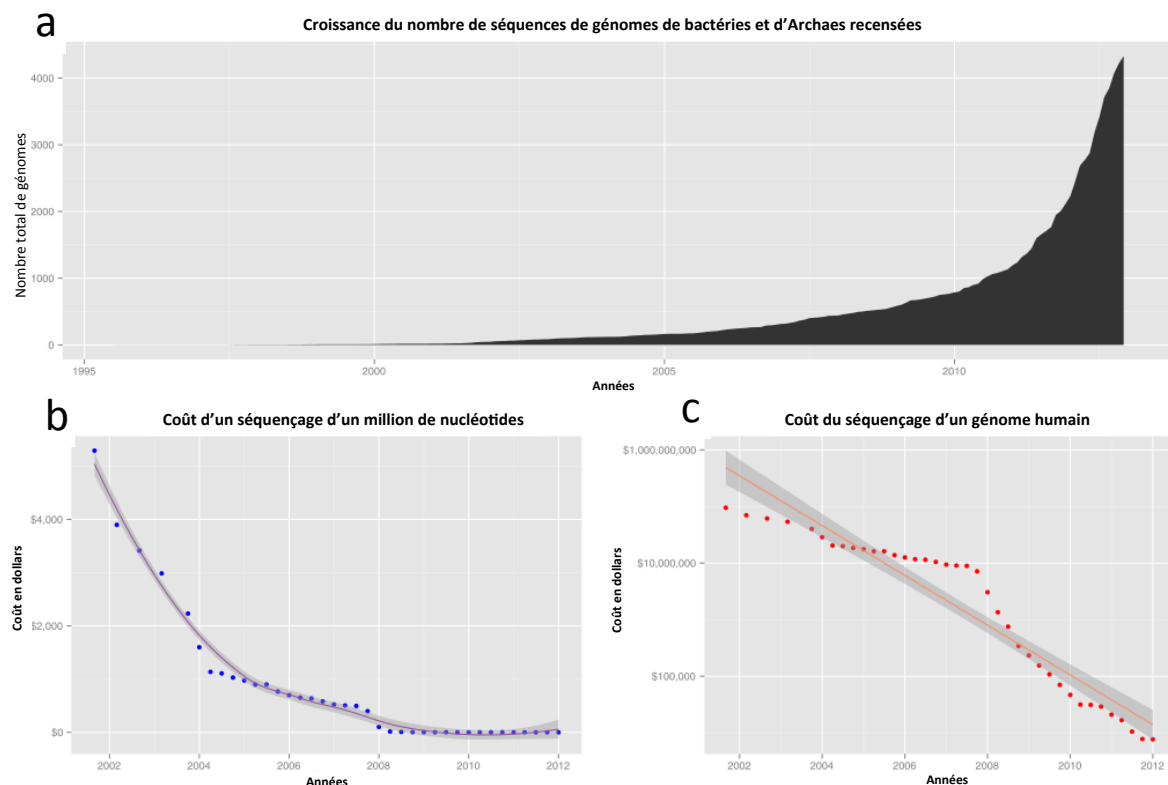


Figure 9. Progrès du séquençage et évolution de son coût

Le nombre de projets en génomique augmente de manière considérable et concorde avec l'amélioration des techniques et la diminution des coûts de séquençages. **a)** Croissance exponentielle des séquences génomiques de bactéries et d'Archaea recensées dans les bases de données depuis 1995. **b)** Le coût du séquençage pour 1 million de nucléotides a nettement diminué depuis 2002, passant de 5,500\$ à 50\$ dix ans plus tard, soit près d'un facteur 100. **c)** Coût du séquençage de 3,000 Mb, soit la taille d'un génome humain en échelle logarithmique, passant de 100,000,000\$ en 2002 à 8,000\$ environ en 2012. Adapté de (Estevezj, 2012) commons.wikipedia.org.

II- Diversité des ET dans le vivant

Le monde du vivant est divisé en trois grands domaines: les Archées, les Eubactéries et les Eucaryotes. L'absence de noyau est une différence morphologique majeure entre procaryotes et eucaryotes. La présence de génomes cytoplasmiques et nucléaires, est également une différence essentielle concernant leurs modes d'évolutions. En effet, lorsqu'un nouvel ET pénètre dans la cellule, ce dernier est directement en contact avec le génome bactérien tandis qu'il est confiné dans le cytoplasme dans le cas d'une cellule eucaryote. On s'attend donc à ce que les trajectoires évolutives des ET et la nature des interactions hôtes-ET soient hautement divergentes entre ces grands groupes.

1- L'architecture des génomes

Les chromosomes sont principalement linéaires chez les eucaryotes et circulaires chez les procaryotes. Les chromosomes circulaires eucaryotes constituent des formes instables, témoins d'anomalies génétiques (chromosomes en anneaux par exemple) ou bien obtenus par ingénierie. Par ailleurs, si certaines lignées bactériennes peuvent contenir de manière sporadique des chromosomes ou plasmides linéaires, cet événement est très rarement observé (*e.g.* les espèces du genre *Caulobacter*).

La taille des génomes procaryotes (10^6 - 10^7 pb) reflète assez bien le contenu en gènes dans le génome, comme illustré par la corrélation positive entre taille de génome et nombre de gènes (Figure 10, Figure 11a) (Koonin, 2011). En effet, les génomes procaryotes comportent une grande densité de gènes, de très courtes régions intergéniques et peu de répétitions. Les rares introns présents dans ces génomes sont souvent de grandes tailles et codent pour des éléments génétiques égoïstes (Toro, 2003). De manière globale, le contenu des génomes procaryotes semble extrêmement dynamique. En effet, certaines séquences génomiques sont parfois partagées par des espèces très éloignées, bien qu'absentes chez les espèces les plus proches. Certains auteurs proposent même que l'évolution de ces génomes seraient majoritairement dominée par des transferts horizontaux (Koonin and Wolf, 2008). Dans ces espèces, les transferts horizontaux de gènes sont largement favorisés par le «mobilome», c'est-à-dire la collection de gènes extra-chromosomiques contenus dans les plasmides et les phages présents de manière transitoire dans les génomes. Comme prédit par les modèles, il a été montré que les génomes procaryotes possédaient en moyenne peu d'ET actifs et que ces derniers étaient finement réprimés par le génome hôte, sans quoi, l'accumulation de mutations pouvait conduire à une perte de fitness, à la maladaptation et un taux de mortalité élevé (Bentley and Parkhill, 2004). De plus, contrairement aux eucaryotes, les génomes procaryotes ont des taux de délétions d'ADN très largement supérieures aux taux

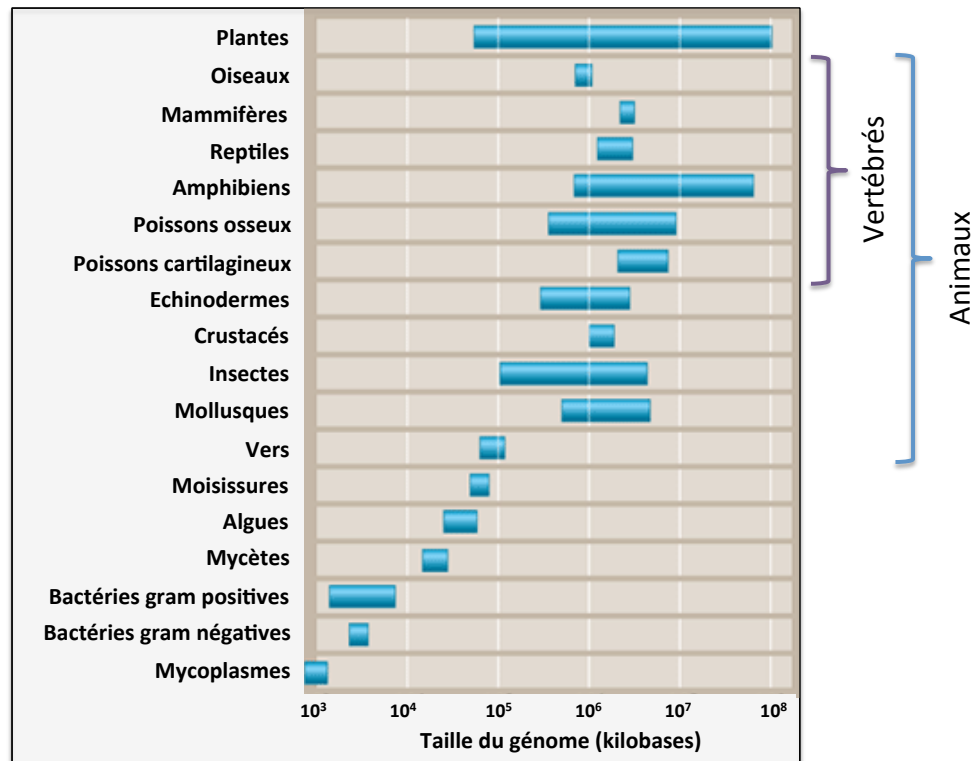


Figure 10. Le paradoxe de la valeur C dans différents organismes

La valeur C (C-value) reflète la taille de génome mesurée en kilobases d'ADN. Intuitivement, on imaginait que la complexité (*e.g.* le nombre de fonctions cellulaires) des organismes était corrélée positivement à la taille du génome. Le paradoxe de la valeur C montre, au contraire, que la taille des génomes n'augmente pas uniformément avec la complexité perçue de l'organisme. Par exemple, quelques amphibiens ont dix fois plus d'ADN que les mammifères, dont l'humain. Malgré ces différences de tailles, le nombre de gènes reste assez similaire chez les vertébrés. Adapté de Transposable Elements, Epigenetics, and Genome Evolution Nina V. Fedoroff.

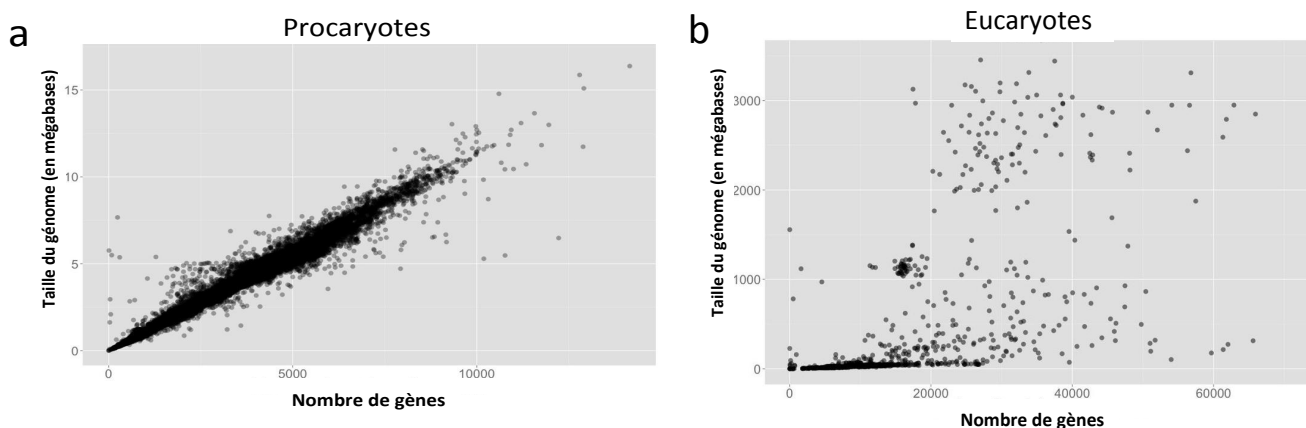


Figure 11. Corrélation entre la taille du génome et le nombre de gènes chez les procaryotes et chez les eucaryotes

Chez les procaryotes (a), l'augmentation de la taille des génomes corrèle bien avec le nombre de gènes codants. Au contraire, chez les eucaryotes (b), aucune corrélation n'est observée. Cela s'explique par la très grande proportion de gènes non codants dont une grande partie inclue des ET. Issu de <https://gist.github.com/roblanf/849789978f6f00358078>. Pour le script, Robert Lanfear from NCBI Data.

d'insertions, ce qui corrobore l'observation d'ET fréquemment soumis à de rapides « turnovers » (Kuo and Ochman, 2009).

Chez les eucaryotes, la taille du génome (10^9 - 10^{11} pb) est très peu corrélée au nombre de gènes (Figure 10, Figure 11b) (Koonin, 2011). Le contenu en ADN non codant, qui correspond à l'ensemble des séquences du génome non traduites ou sans fonctions biologiques identifiées, est extrêmement variable et représente une grande portion de ces génomes. Ces séquences non codantes sont retrouvées principalement dans les régions intergéniques, bien que les séquences codantes soient généralement interrompues par des introns non codants. Contrairement à ce qui observé chez les procaryotes, cette large fraction d'ADN non codante provient rarement de fragments d'ADN aléatoires transférés horizontalement depuis l'environnement cytoplasmique. En effet, le noyau représente une barrière physique limitant la fréquence des transferts horizontaux (Lechardeur and Lukacs, 2006; Pollard et al., 2001; Shimizu et al., 2005). Certains virus outrepassent néanmoins régulièrement cette barrière, c'est le cas par exemple des rétrovirus à ARN. Ces derniers, qui se répliquent uniquement dans le noyau, produisent un ADN complémentaire capable de s'intégrer dans le génome hôte. Ainsi, les infections par diverses familles de rétrovirus constituent un facteur majeur de l'évolution des génomes eucaryotes (Holmes, 2011). De manière générale, ces rétroéléments et les répétitions d'ADN associées, constituent une large fraction de la quantité d'ADN non codante chez les eucaryotes. Il a été proposé par ailleurs, que ce type de répétitions était régulièrement impliqué dans l'évolution des caryotypes, en facilitant l'échange, la fusion et la séparation de bras chromosomiques au niveau des centromères et des télomères (Eichler and Sankoff, 2003). De tels remaniements sont observables pour des organismes phylogénétiquement proches (*e.g.* l'homme et la souris) qui possèdent un nombre de chromosomes et une architecture génomique très distinctes malgré un contenu en gènes similaire. Les chromosomes linéaires eucaryotes diffèrent de leurs homologues circulaires procaryotes, notamment *via* la présence de centromères et de télomères. Ce type de régions hautement répétées est associé à la présence de protéines spécifiquement impliquées dans la compaction de l'ADN *via* les nucléosomes. Cette spécificité eucaryote est essentielle au contrôle de l'activité des séquences répétées et à la gestion de la quantité massive d'ADN non codante.

Ces exemples illustrent la nature divergente des interactions hôtes/éléments répétées entre ces grands clades. Les turnovers rapides de contenu en séquences répétées chez les procaryotes contrastent fortement avec la tendance à une forte accumulation chez les eucaryotes, pour lesquels des mécanismes dédiés à leur gestion ont rapidement été sélectionnés (Zeh et al., 2009).

2- ET et procaryotes

L'absence de reproduction sexuée chez les procaryotes, c'est-à-dire l'absence de recombinaison entre allèles paternels et maternels, constitue une spécificité majeure et déterminante quant à leur mode d'évolution. Cependant, un certain nombre d'espèces eucaryotes se reproduisent de manière clonale et il est intéressant de les comparer aux autres espèces procaryotes pour en déduire les effets globaux du sexe sur l'évolution des génomes.

Chez une population asexuée, il est attendu qu'une prolifération non contrôlée des ET conduise rapidement à son extinction. En effet, l'absence de recombinaison entre chromosomes entraîne l'accumulation d'ET par cliquet de Muller ((Muller, 1964) «*relation of recombination*»). A termes, les lignées accumuleraient un grand nombre de mutations délétères et seraient rapidement contre-sélectionnées en faveur de lignées accumulant des formes domestiquées d'ET ou bien des éléments inactifs. Ce phénomène a été observé chez les rotifères, une espèce eucaryote ayant perdu la capacité de se reproduire de manière sexuée (Arkhipova and Meselson, 2000). Par ailleurs, les modèles théoriques prédisent que les échanges génétiques chez les espèces sexuées diploïdes permettent de minimiser les effets délétères induits par des événements de transpositions, ce qui en expliquerait partiellement le maintien (*e.g.* la perte d'insertions hétérozygotes au cours du brassage aléatoire des chromosomes au cours de la méiose) (Dolgin and Charlesworth, 2008). Etant donné la fréquence des transferts horizontaux chez les bactéries, certains auteurs n'hésitent pas à parler de sexe, dans la mesure où des processus d'échanges de matériels génétiques existent. Il s'agit d'une "sexualité" unidirectionnelle qui ne permet pas un réarrangement massif de toutes les combinaisons alléliques possibles (en opposition avec le sexe méiotique, où des échanges bidirectionnels ont lieu entre deux individus). Il s'avère néanmoins que ce type d'échange génétique est très fréquent chez les bactéries et met en jeu des mécanismes qui leurs sont propres.

Les éléments de type *IS* (*Insertion Sequence*) sont très largement répandus dans les réplicons procaryotes (Figure 12a). L'analyse génomique de leur distribution montre une faible diversité intragénomique qui contraste fortement avec la forte divergence entre génomes au sein d'une espèce. Lorsque l'on mesure le taux d'évolution des *IS* (*ratio* de substitutions non-synonymes sur synonymes), ces derniers sont étonnamment faibles relativement aux gènes dupliqués. Une analyse effectuée chez 376 génomes suggère que les invasions de l'hôte par les *IS* sont récentes (Wagner, 2006). L'interprétation donnée par Wagner est que le maintien des *IS* décrit des cycles d'extinctions et de réinvasions périodiques entre les différentes lignées bactériennes. Dans certains cas, le génome d'une espèce contient des centaines de familles d'*IS* distinctes tandis que l'espèce sœur n'en contient aucune, suggérant que les ET sont soumis à des dynamiques particulièrement rapides. Globalement, les phylogénies des espèces ne sont pas congruentes avec les phylogénies des *IS*, ce qui corrobore l'hypothèse des transferts horizontaux fréquents (Figure 12b).

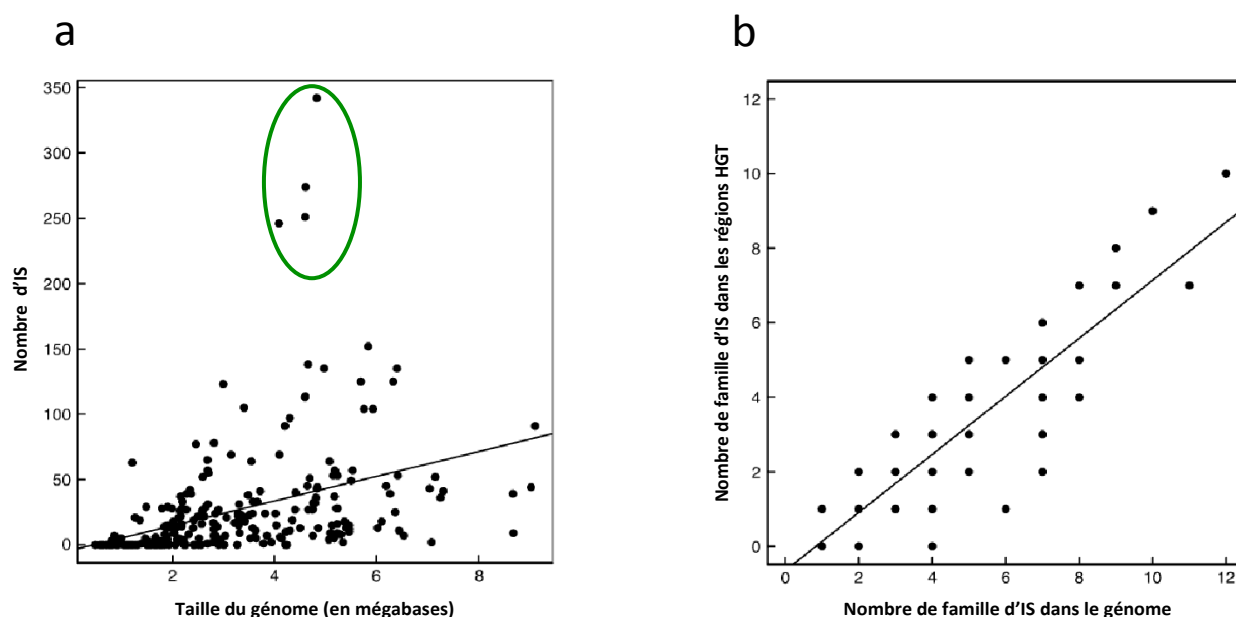


Figure 12. Nombre d'IS (séquence d'insertion) présentes dans le génome en fonction de la taille du génome

La figure (a) montre que l'augmentation de la taille des génomes est corrélée à la quantité d'IS (séquence d'insertion) présentes après ré-annotation de 262 génomes procaryotes (Touchon et Rocha, 2007). On note cependant que certaines lignées possèdent un très grand nombre d'IS sans que la taille du génome soit très affectée (cercle vert). La Figure (b) montre le contenu en IS dans les régions associées à un enrichissement en transferts horizontaux (ordonnées) en fonction du nombre de familles d'IS présentes dans le génome (abscisses). La densité en IS dans les régions riches en transferts horizontaux est supérieure relativement au reste du génome, on y retrouve de plus, la même diversité de famille indiquant que le contenu en IS dans le génome est principalement issu de transferts horizontaux récents.

Un taux élevé de transferts horizontaux semble donc crucial, d'une part pour expliquer la présence d'une même IS dans des espèces éloignées alors qu'elle est absente dans les espèces proches, d'autre part pour compenser cette forte tendance à l'extinction (Touchon and Rocha, 2007). Cela reste néanmoins une hypothèse soutenue uniquement par la non-congruence des phylogénies, mais à ce jour, non validée par l'expérience. Aujourd'hui encore, on ignore si les ET procaryotes sont régulièrement éliminés par la sélection purifiante qui lutte contre les effets délétères associés à leur accumulation, ou bien si leurs transferts très fréquents, favorisent une forme de « sexualité » garante de la variabilité génomique nécessaire à l'adaptation des lignées.

De nouvelles perspectives ont été apportées par la découverte du système CRISPR, un mécanisme dédié à la répression des éléments mobiles. Il constitue, au même titre que le système des *Piwi interacting RNA* chez les eucaryotes, un mécanisme de défense contre les séquences nucléiques exogènes et passe par la production de petits ARN interférents. Les propriétés d'un tel mécanisme commencent à être étudiées du point de vue de la régulation des ET et semblent avoir un rôle essentiel pour limiter l'expansion des IS et la diversification des lignées induites par recombinaison entre IS (Watanabe et al., 2013).

3- Diversité des ET chez les eucaryotes

Les génomes eucaryotes sont caractérisés par leur grande taille, dont les variations recouvrent plusieurs ordres de grandeurs (Figure 10). Parmi les facteurs déterminants de l'évolution de leur taille, le contenu en ADN non codants (composé en grande partie des ET) en est le contributeur majoritaire (Figure 13). Au niveau évolutif, la question est de savoir si l'augmentation de la taille des génomes a un rôle adaptatif, s'il s'agit d'une conséquence de l'activité égoïste des éléments transposables (“*selfish DNA*”, (Doolittle and Sapienza, 1980; Orgel and Crick, 1980)), ou tout simplement d'une augmentation neutre liée à des effets de dérives (Lynch, 2007). Il est cependant clair que les variations de tailles de génomes sont globalement corrélées au contenu en éléments transposables (Kidwell, 2002). Les plus gros génomes tendent en effet, à accumuler une plus grande quantité de transposons et de rétrotransposons dans les régions pauvres en gènes (Gregory, 2004; Vinogradov, 2005).

A - Quelques exemples extrêmes

Chez les eucaryotes unicellulaires, la plupart des espèces séquencées sont des pathogènes et des parasites. Cependant, l'abondance et la diversité des ET présents sont extrêmement variables. Le génome du parasite *Trichomonas vaginalis* (160 Mb), contient des séquences répétées à hauteur de 75%, dont 25% sont des éléments mobiles (Carlton et al., 2007). Ces séquences recouvrent une large diversité d'éléments répétés, qui sont, soit d'origine virale, soit des rétrotransposons soit des transposons à ADN. Chacune de ces trois grandes familles est présente en petit nombre (< 1000 copies) et a la particularité de présenter des copies très semblables en terme d'identité de séquences, suggérant des “bursts” de transposition récents. Le génome du parasite causant la malaria (*Plasmodium falciparum*) ne présente quant à lui aucun ET (Gardner et al., 2002). De même, les génomes de *Trypanosoma* et de la *Leishmania*, ne possèdent qu'entre 2 et 5% d'ET qui sont uniquement des rétrotransposons (Bringaud et al., 2008). Pour les espèces appartenant au genre *Entamoeba*, la fraction en éléments transposables varie du simple au double (10% chez *E. dispar* et 20% chez *E. histolytica*) (Feschotte and Pritham, 2007; Lorenzi et al., 2008). Plus étonnant encore, certaines espèces ne contiennent quasi-exclusivement que des rétrotransposons non-LTR (*E. histolytica* et *E. dispar*), tandis que d'autres, contiennent très majoritairement des transposons à ADN (*E. invadens*).

Le cas le plus extrême concernant l'amplification des ET est observé chez les plantes. Plus de 80% du génome du maïs est constitué d'ET, composé en majorité des éléments *copia* et *gypsy* qui correspondent respectivement à 24 et 46% de la taille du génome (Schnable et al., 2009). Le genre *Oryza* (le riz domestiqué) est un exemple frappant en termes de variations de taille. En effet, le génome d'*Oryza Australiensis* a doublé en l'espace de 3 millions d'années, en dépit d'événements de polyploïdisation. L'amplification de plusieurs familles de rétrotransposons à LTR explique majoritairement ces variations de tailles (Piegu et al., 2006). Ces observations contrastent fortement avec d'autres espèces modèles, telle qu'*Arabidopsis thaliana*, dont le contenu

en ET ne représente que 10% du génome (Arabidopsis Genome, 2000). Ces exemples, parfois extrêmes, illustrent la variabilité des dynamiques d'ET. Les exemples les plus marquants sont néanmoins souvent associés au très large succès d'une ou deux familles exclusivement, qui sont associées à de fortes augmentations de tailles du génome. Chez les animaux, on retrouve des exemples similaires, c'est le cas, par exemple, des L1 chez les mammifères que nous discuterons en détail ci-après.

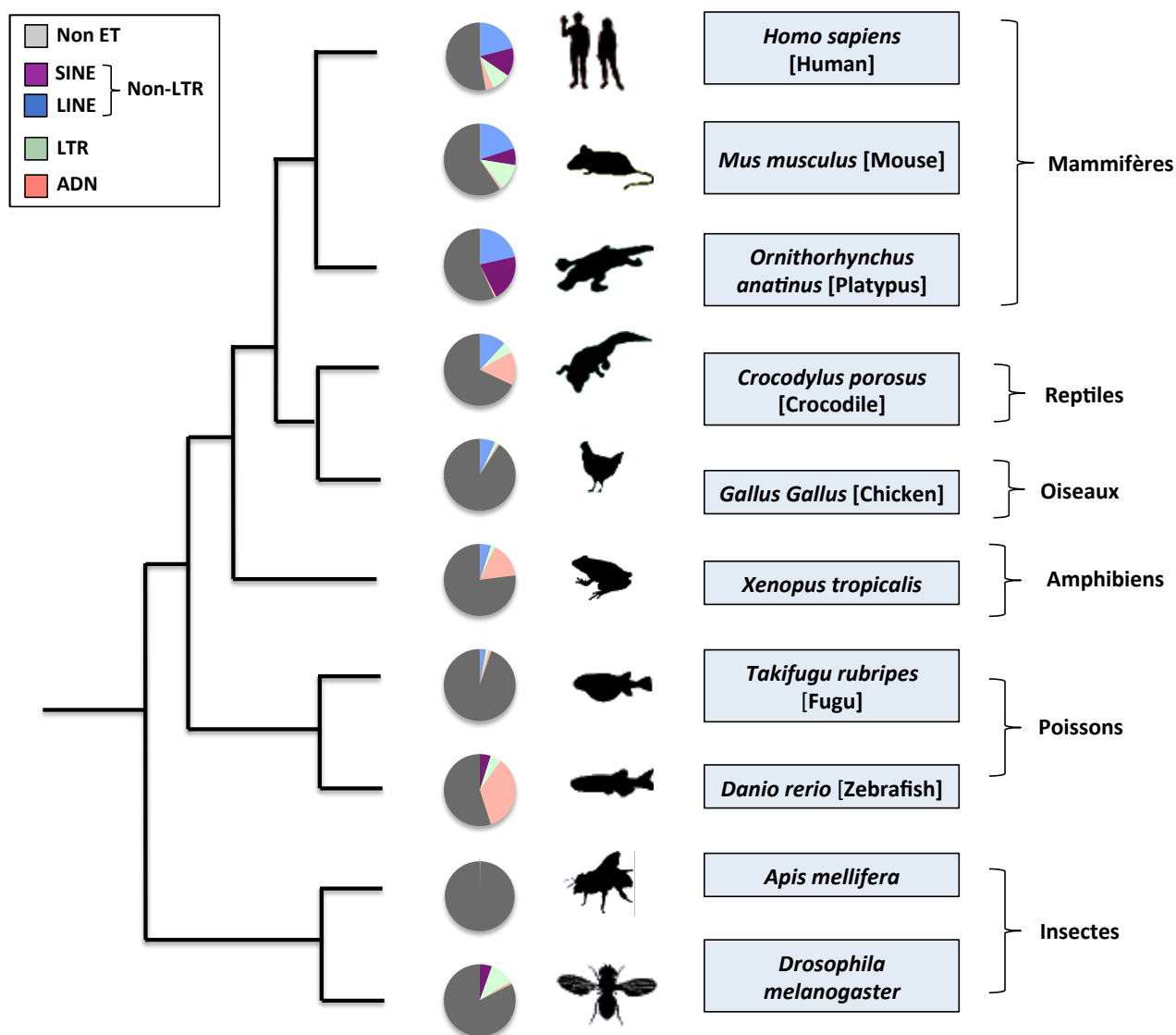


Figure 13. Représentation de la diversité des ET chez les métazoaires

La figure montre les fractions de génomes d'animaux occupées par les grandes familles d'ET. Cette figure résume les variations considérables de contenu en ET entre différents clades. La distribution des ET dans ces génomes semble stochastique et souvent liée à l'amplification massive d'une seule famille dans un clade (e.g. les LINE chez les mammifères – bleu). On note aussi, chez certaines espèces, la perte quasi-totale d'ET (comme chez l'abeille), souvent associée à des phénomènes de contraction de génome et à la perte de grande quantité d'ADN.

B - Les ET chez les Arthropodes

Chez les insectes, l'espèce la mieux caractérisée est *D. melanogaster* (diptère) (Figure 14), chez laquelle les ET représentent environ 18% du génome et moins de 4 % de la fraction euchromatique (Kaminker et al., 2002), contrairement à son espèce sœur, *D. simulans*, qui ne compte que 5% d'ET. Chez ces deux espèces, les ET sont essentiellement hétérochromatiques et dominés à 60% par les rétrotransposons à LTR. Chez le *Tribolium* rouge de farine (coléoptère), les 6% d'ET présents dans le génome montrent une distribution plus équilibrée et compte un grand nombre de familles diverses (LTR, sans LTR et transposons à ADN) (Tribolium Genome Sequencing et al., 2008). Chez les lépidoptères, le genre *Bombyx* compte au moins 35% de séquences dérivées d'ET. Chez l'espèce *Bombyx mori*, plus de 50% des ET appartiennent à la famille des *gypsy* et sont associés à une vague d'amplification rapide au cours des 5 derniers millions d'années (Osanai et al., 2006). L'exemple le plus étonnant concerne le génome de l'abeille, *Apis mellifera*. Cette dernière possède moins d'1% d'ET et cette fraction est essentiellement composée d'éléments de type *mariner* anciens et dégradés (Honeybee Genome Sequencing, 2006).

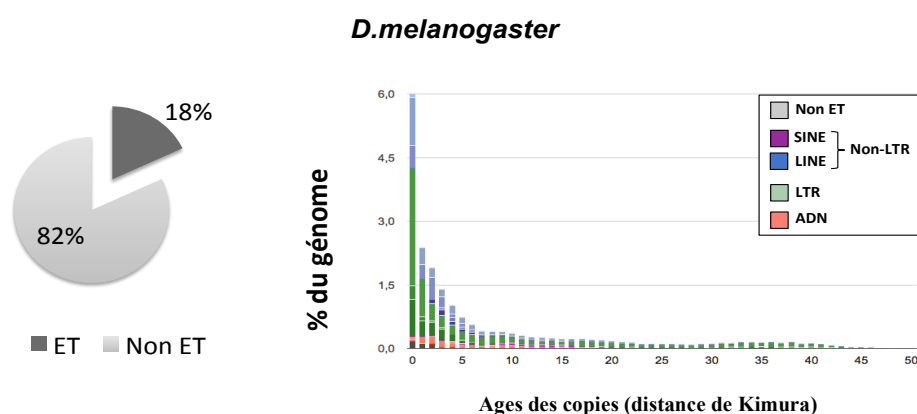


Figure 14. Répartition et proportion en ET chez *D. melanogaster*

Le panneau de gauche montre le contenu en ET chez *D. melanogaster*, exprimé en pourcentage de la taille totale du génome. Environ 18% du génome est constitué d'ET chez *D. melanogaster*. Le panneau de droite montre la fraction du génome (ordonnées) en ET en fonction de l'âge des copies (abscisses) chez *D. melanogaster*. On observe une majorité de copies dont l'amplification est massive et récente. Les ET sont dominés à 60% par des rétrotransposons à LTR (e.g. *Gypsy*), suivi de rétrotransposons sans LTR (e.g. *Jockey*).

C - Les ET chez les vertébrés

Chez les vertébrés, la co-évolution des ET et des génomes hôtes montre aussi des différences très marquées au sein des clades et entre clades. Les analyses comparées des grands groupes de vertébrés ont révélé que les populations de rétrotransposons et de transposons à ADN sont quantitativement et qualitativement très variables. En effet, on retrouve souvent des différences importantes au sein de lignées génomiques très proches (Bohne et al. 2008). Cela est essentiellement due 1) à l'évolution d'ET ancestraux dont les *scenarios* évolutifs s'étendent de l'extinction totale à l'invasion massive des familles, 2) à l'introduction de nouvelles lignées issues

d'infections ou de transferts horizontaux spécifiques d'une lignée (e.g. les rétrovirus endogènes), 3) à l'émergence de nouveaux ET, comme on peut l'observer pour les rétroéléments non codants de type *SINE*. On suppose à l'heure actuelle, que les ET constituent une force majeure capable d'initier d'importantes transitions évolutives et favorisant la divergence entre lignées génomiques proches (Bohne *et al.* 2008).

Les poissons (téléostéens) représentent, en termes de nombres d'espèces (environ 30 000), une diversité plus importante que l'ensemble des autres groupes de vertébrés réunis (mammifères, amphibiens, reptiles et oiseaux). Comparé aux mammifères, les génomes de poissons sont plutôt compacts et de petite taille. Par exemple, le poisson lune (pufferfish) a un génome de l'ordre de 400 Mb, parmi les plus petits génomes de vertébrés (environ 10 fois plus petit que le génome humain et seulement 3 fois plus gros que celui de *D. melanogaster*). Cette différence s'explique en grande partie par le faible contenu en ET dans le génome (Volff, 2005). Parmi les espèces séquencées, on retrouve des exemples similaires chez *Takifugu rubripes* (Figure 15) et *Tetradon nigroviridis*, qui comptent moins d'1% d'ET dans le génome (Volff *et al.*, 2003). De manière intéressante, on y retrouve toutes les grandes familles d'ET (LTR, sans LTR, transposons à ADN) avec une large diversité au sein des sous-familles d'éléments.

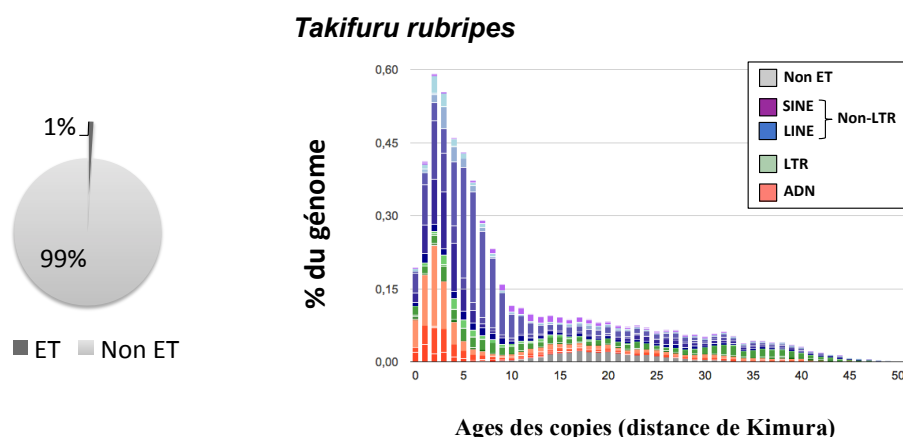


Figure 15. Répartition et proportion en ET chez *Takifuru rubripes*

Le panneau de gauche montre le contenu en ET chez *Takifuru rubripes*, exprimé en pourcentage de la taille totale du génome. Moins d'1% du génome est constitué d'ET, l'un des plus faibles pourcentages parmi les vertébrés. Le panneau de droite montre la fraction du génome (ordonnées) en ET en fonction de l'âge des copies (abscisses) chez *Takifuru rubripes*. On observe un burst d'amplification récent de transposons à ADN (orange) précédé d'une vague d'amplification d'ET de type *LINE* (bleu).

Cependant, bien que cette diversité représente un grand nombre de familles différentes, toutes sont présentes en très faible nombre de copies. Chez le poisson zèbre (*Danio rerio*), on retrouve aussi une grande diversité de familles d'ET. Cependant, son génome est quatre fois plus gros que celui du poisson lune (1700 Mb) et largement associé à l'amplification massive et relativement ancienne de transposons à ADN (environ 30 % de la taille totale du génome). En outre, le clade des rétrotransposons sans LTR de type *LINE* constitue plus de 30 lignées distinctes très apparentées. Il semble, chez cette espèce, que le génome a subi de nouvelles vagues

d'invasions par différentes familles de *LINE*, qui font suite aux vagues plus anciennes des transposons à ADN. La diversité des *LINE*, commune à plusieurs espèces de poissons, contraste fortement avec le cas des mammifères où une seule grande lignée de *LINE* se retrouve très largement représentée. Chez les téléostéens, on retrouve de manière générale une distribution bimodale caractérisée par un ensemble de copies très anciennes (très divergentes), qui coexistent avec des lignées en cours de réinvasion (très apparentées). En termes de dynamique d'invasion, ces observations suggèrent un renouvellement rapide des lignées qui s'amplifient massivement et régulièrement sur de courtes périodes de temps (Duvernell et al., 2004; Furano et al., 2004). Une des hypothèses évoquée pour expliquer ce type de dynamique est associée à la perte récurrente d'ADN chez les poissons (Blass et al., 2012). Cela corrobore les variations de tailles de génome entre espèces ainsi que la perte récurrente et massive de plusieurs familles d'ET. L'exemple du rétroélément *Rex1* est assez explicite. Il est caractérisé par une distribution discontinue chez tous les Actynoptériens. On le retrouve quasiment éteint chez certaines espèces et amplifié à plus de 5000 copies chez d'autres (Volff et al., 2000).

Le génome de l'amphibien (*Xenopus tropicalis*) est, quant à lui, largement dominé par des transposons à ADN qui représentent 25% de la taille du génome (Figure 16), contre 9% seulement pour les éléments de Classe I. Cinq familles majeures de transposons à ADN représentent à eux seuls environ 20 % du génome (*Kolobok*, *hAT*, *Harbinger*, *mariner* et *piggyBac*) et on retrouve aussi un grand nombre de copies de la famille des *hélitrons* et *polintons*. Cependant, bien que présents en petits nombre de copies, les transposons à LTR recouvrent une plus grande diversité de familles que celle retrouvée dans l'ensemble des autres espèces de vertébrés, suggérant leur origine très ancienne. Par ailleurs, parmi les familles les plus actives récemment, on retrouve une fois encore des rétrotransposons de type *LINE/CRI* (Hellsten et al., 2010).

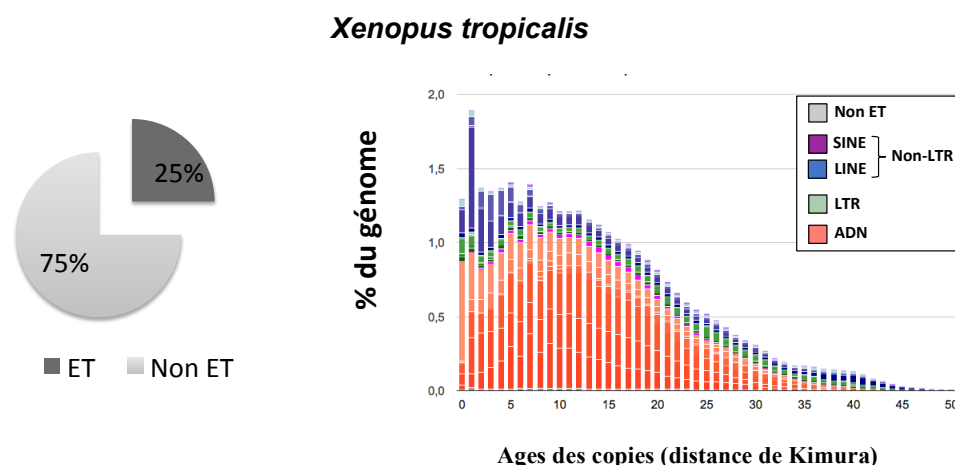


Figure 16. Répartition et proportion en ET chez *Xenopus tropicalis*

Le panneau de gauche montre le contenu en ET chez *Xenopus tropicalis*, exprimé en pourcentage de la taille totale du génome. Environ 25% du génome est constitué d'ET, majoritairement des transposons à ADN (20% du génome). Cette espèce est l'une des rares espèces de vertébrés séquencées ayant une forte proportion de transposons à ADN. Le panneau de droite montre la fraction du génome (ordonnées) en ET en fonction de l'âge des copies (abscisses) chez *Xenopus tropicalis*. On observe un burst relativement ancien de transposons à ADN (orange), suivi de deux vagues d'amplification récentes d'ET de type *LINE* et d'autres familles de transposon à ADN.

Chez les reptiles, le génome du lézard *Anolis carolinensis* a été récemment séquencé (Alfoldi et al., 2011). Il s'agit d'un génome relativement petit (1 700 Mb) assez équilibré concernant la présence des différentes familles d'ET. Les 35 % du génome associés aux ET, contiennent une très large diversité de rétrotransposons de type *LINE* (*L1*, *L2*, *CR1*, *RTE* et *R4*), caractérisé par la diversification récente de plusieurs familles. Par exemple, les 20 familles de *L1* et les 17 familles de *L2* ont un niveau de divergence inférieur à 2%. Les familles de transposons à ADN (majoritairement des *mariner* et des *hAT*) et les rétrotransposons à LTR (*Metaviridae* et *Pseudoviridae*), sont quant à eux issus de vagues d'amplifications plus anciennes. D'une manière générale, les squamates (lézards et serpents) montrent des niveaux élevés de variations en termes d'ADN répétés. En effet, les comparaisons de caryotypes et de séquençages incomplets entre espèces, ont révélé que les séquences répétées ne s'établissent pas sur le long terme et sont caractérisées par leur grande hétérogénéité. De manière intéressante, on peut associer ces grandes variations de contenu en séquences répétées à une incroyable diversité caryotypique (Hall, 1973a; Hall, 1973b). Au sein d'espèces proches, ou même entre populations de la même espèce, le nombre et la morphologie des chromosomes peuvent varier considérablement (e.g. le complexe de populations chez l'espèce *Lacerta kulzeri*).

Cette hétérogénéité du génome chez les squamates contraste fortement avec l'étonnante stabilité des génomes de crocodiles et de tortues. Chez la tortue par exemple, la structure hétérochromatinienne du chromosome 1, visualisée en cytologie, ne montre aucune différence entre les espèces les plus divergentes (Bickham, 1981). Les génomes de tortues et crocodiles sont beaucoup plus gros que ceux des squamates et sont organisés en blocs d'hétérochromatine extrêmement bien conservés. Ces régions sont pauvres en gènes (Mouchiroud et al., 1991) et possèdent de longs introns enrichis en séquences répétées (Vinogradov, 1999). L'absence de recombinaison dans ces régions, envahies par du « junk DNA », semble avoir figé l'état du génome depuis près de 200 millions d'années. Cette stase structurale a récemment été confirmée par le séquençage de trois génomes crocodiliens. Parmi les ET annotés, 95% appartiennent à des familles présentes dans les trois génomes, à peu près aux mêmes fréquences. Seulement 5% des copies (< 2% du génome) sont apparues après la divergence entre les Longirostres (crocodiles et gavials) et les alligators, il y a environ 100 millions d'années. Ces résultats suggèrent que les succès d'invasions de nouvelles familles d'ET ont globalement diminués à l'exception d'un « burst » mineur chez l'ancêtre commun des Longirostres. Lorsqu'on estime le nombre de familles actives depuis la divergence mammifères-crocodiles (235 millions d'années), on compte un total de 823 familles, soit un taux d'environ 3,5 nouvelles familles d'ET par million d'années). Ce taux est cependant inférieur à 1 chez les crocodiles et les gavials depuis leur divergence. De plus, si on s'intéresse aux familles d'ET très anciennes, on les retrouve surreprésentées chez les crocodiliens, relativement aux autres espèces. Parmi les 74 familles d'ET présentes depuis plus de 300 millions d'années (avant la divergence entre reptiles et mammifères), une grande quantité d'ADN dérivée de ces éléments ancestraux est retrouvée chez les crocodiles. Chez les crocodiliens la quantité d'ET ancestraux est 5 fois supérieure à ce qui est observé chez les mammifères, 3,8 fois supérieure comparé au poulet et 15 fois supérieure à celui d'*Anolis carolinensis*. Seul le génome de la tortue montre un taux plus élevé d'ADN dérivés d'ET anciens, suggérant une stase génomique plus importante

encore. Ces ratios donnent une estimation qualitative des différences de taux de mutations et délétions dans les génomes. Au vu de ces données, il semble que la vitesse d'évolution du génome des crocodiles et des tortues soit exceptionnellement lente.

Chez les oiseaux, le génome du poulet ne montre aucun signe d'activité récente d'ET (International Chicken Genome Sequencing, 2004) (Figure 17). La présence de plusieurs lignées divergentes de rétrotransposons de type *LINE*, suggère que ces éléments étaient actifs dans le passé. Les ET annotés, représentent environ 10% de la taille du génome, une fraction faible comparée aux mammifères (34 à 52%). Le séquençage de 48 nouvelles espèces aviaires tend à confirmer cette tendance. En effet, les génomes contiennent généralement peu d'éléments répétés (entre 4 et 10%), ce qui en fait le taux le plus faible parmi les tétrapodes. La seule exception est le pic mineur observé, correspondant à *Picoides pubescens*, qui contient 22% d'ET qui dérivent pour la plupart d'une expansion spécifique de *LINE* (de type CR1). Par ailleurs, il est intéressant de constater que la fraction génomique des éléments de type *SINE*, est de 10 à 30 fois inférieure en comparaison aux reptiles, suggérant que cette perte massive a eu lieu chez l'ancêtre commun des oiseaux. De manière intéressante, les caryotypes sont extrêmement stables chez les oiseaux et certains auteurs associent cette stabilité au faible contenu en ET (Ellegren, 2010; Pardo-Manuel de Villena and Sapienza, 2001). Plusieurs exemples montrent que leurs chromosomes sont restés intacts pendant près de 100 millions d'années (Griffin et al., 2007). Par ailleurs, la majorité des espèces possèdent entre 76 et 80 paires de chromosomes dont la synténie est extrêmement bien conservée (Nanda et al., 2008). Enfin, les taux étonnamment faibles de duplications segmentales et des réarrangements chromosomiques ont eux aussi été associés au faible contenu en séquences répétées (Griffin et al., 2008).

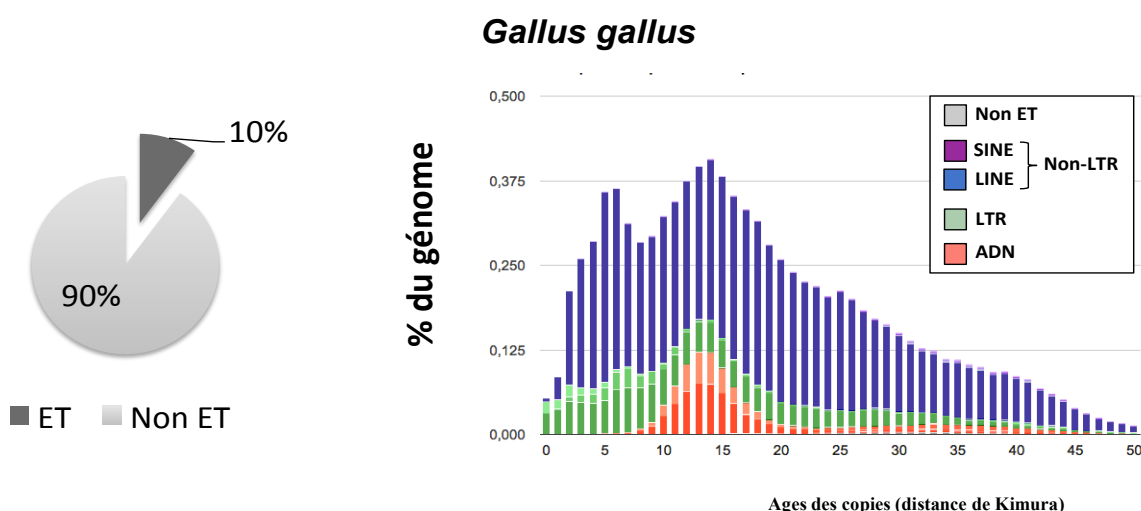


Figure 17. Répartition et proportion en ET chez *Gallus gallus*

Le panneau de gauche montre le contenu en ET chez *Gallus gallus*, exprimé en pourcentage de la taille totale du génome. Seulement 10% du génome est constitué d'ET, très largement dominé par les rétrotransposons de type *LINE* (CR1). Chez les oiseaux, le contenu en ET est globalement faible et associé à des tailles de génomes plus petites relativement aux autres vertébrés. Le panneau de droite montre la fraction du génome (ordonnées) en ET en fonction de l'âge des copies (abscisses) chez le poulet (*Gallus gallus*). On observe une amplification continue d'éléments de type *LINE* (bleu) d'origine très ancienne, qui comprend un burst plus récent. On note aussi une vague d'amplification ancienne de transposons à ADN (orange) dont la majorité des familles semble éteinte aujourd'hui.

Concernant d'autres propriétés du génome, la taille moyenne des gènes codant les protéines, sont deux fois plus courtes que celles des mammifères. Cette réduction de taille est essentiellement due au raccourcissement des introns et des séquences intergéniques et conduit à une forte densité de gènes. Une telle contraction du génome est observée de manière convergente chez la chauve-souris, le seul mammifère volant. Une des hypothèses serait que la contraction du génome pourrait être liée à l'acquisition du vol, qui demande des ressources métaboliques conséquentes qui impliquent une régulation rapide des gènes (Zhang et al., 2013).

Les génomes de mammifères ont une taille de l'ordre de plusieurs milliers de mégabases. Ceux-ci sont largement dominés par des rétrotransposons sans LTR de type *LINE* ou *SINE*. Cela reste vrai pour les monotrèmes (ornyothorinque), les métathériens (marsupiaux) et les euthériens (placentaires). La principale différence entre ces trois groupes, est la présence majoritaire du clade *L1* chez les euthériens et les métathériens, tandis que le clade *L2* est prédominant chez les monotrèmes (Feschotte and Pritham, 2007; Mouse Genome Sequencing et al., 2002; Warren et al., 2008). De manière générale, chez l'ensemble des mammifères, l'activité des *L1* décrit de fortes vagues d'amplification suivies de longues périodes d'inactivité (Furano, 2000). Cependant, ces derniers ont perdu toute trace d'activité chez quelques rares espèces (Cantrell et al., 2008; Casavant et al., 2000). Chez les euthériens, le génome le plus étudié est celui de l'homme, très largement dominé par les *L1* qui représentent au moins 21% du génome, avec sa contrepartie d'éléments non-autonomes, les *Alu* présents à plus d'un million d'exemplaires (Figure 18). L'explosion de ce clade chez l'homme, résulte de l'activité d'une seule lignée active depuis l'origine des euthériens. Bien que la diversité des ET soit limitée aux *L1* et aux *Alu*, cela n'a pas toujours été le cas. En effet, il semble que les transposons à ADN étaient plus diversifiés et actifs chez les primates ancestraux (Feschotte and Pritham, 2007), il y a 37 millions d'années. L'extinction des rétrotransposons à LTR semble être plus lointaine, mais on détecte un grand nombre de rétrovirus endogènes qui représente environ 8 % du génome des primates (Barbulescu et al., 1999).

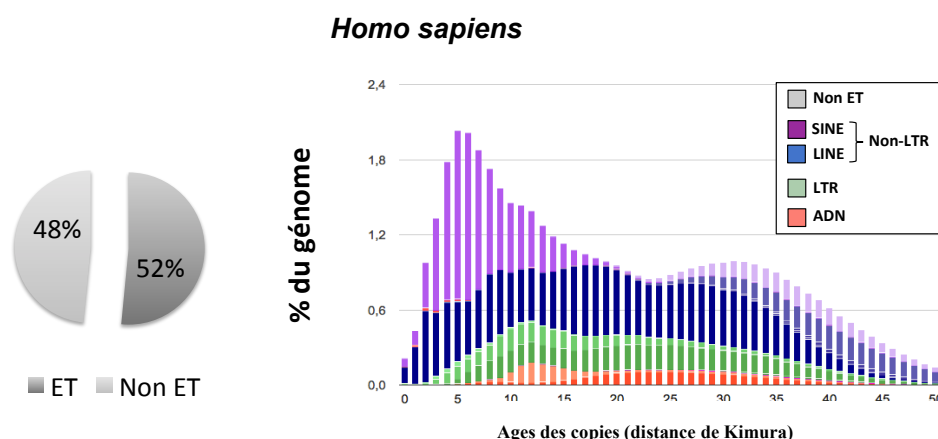


Figure 18. Répartition et proportion en ET chez *Homo sapiens*

Le panneau de gauche montre le contenu en ET chez *Homo sapiens*, exprimé en pourcentage de la taille totale du génome. Plus de 50% du génome est constitué d'ET, très largement dominé par les rétrotransposons de type *LINE* (*L1*). Le panneau de droite montre la fraction du génome (ordonnées) en ET en fonction de l'âge des copies (abscisses) chez l'homme (*Homo sapiens*). On observe une amplification continue d'éléments de type *LINE* (bleu) suivi d'un burst massif et relativement récent d'éléments de type *SINE* (mauve).

De manière générale, les profils d'ET sont semblables dans les autres génomes euthériens, à l'exception de quelques groupes qui dévient significativement du modèle humain. Ces exceptions sont associées à la colonisation récente de rétrovirus actifs dans la lignée germinale. Ceux-ci ont produit un grand nombre de copies qui viennent s'ajouter à la diversité des ET dans ces taxons. Chez les *Oryzomys* (rat du riz) par exemple, les éléments *MysTR* se sont amplifiés à plus de 10.000 copies (Cantrell and Bresnahan, 2005). De même, les éléments de type *IAP* (rétrotransposons) représente le groupe d'ET le plus actifs chez la souris (Ribet et al., 2008). Certains transferts horizontaux d'ET ont eux aussi contribué à l'expansion de certaines familles au sein de lignées uniques. C'est le cas de plusieurs types de transposons à ADN et d'éléments de type RTE (rétrotransposons apparentés à ceux des reptiles) chez la chauve-souris, ou encore chez la vache (Kordis and Gubensek, 1998; Ray et al., 2008).

Au regard de l'extrême diversité des cas présentés (Figure 19), il semble qu'il n'y ait pas de tendance globale pour expliquer le succès, le maintien ou l'extinction des ET dans les génomes eucaryotes. Que ce soit chez les protistes, les plantes, les insectes ou les vertébrés, certaines espèces contiennent un très grand nombre d'ET tandis que d'autres n'en contiennent quasiment pas. Le même constat peut être fait concernant la diversité des éléments présents: certaines espèces contiennent une grande diversité de familles tandis que d'autres ne contiennent majoritairement qu'une lignée d'un clade bien précis. Il ne semble pas non plus y avoir de quelconques corrélations entre le nombre d'innovations fonctionnelles des organismes et le répertoire d'ET présents, suggérant que leur rôle général dans les processus d'adaptations est négligeable. En revanche, la plupart des cas associés à une augmentation rapide de la taille du génome, semble majoritairement liée à l'expansion rapide d'un seul type de transposon (par exemple, les *LI* chez les euthériens ou les rétrotransposons à LTR chez les plantes), plutôt qu'à l'amplification massive de multiples familles.

L'activité des ET semble par ailleurs affecter certaines propriétés du génome. Indépendamment de la quantité de séquences répétées présentes, l'apparente inactivité des ET chez les crocodiles, les tortues et les oiseaux, semble être associée à l'étonnante stase chromosomique qui persiste depuis plusieurs millions d'années. Au contraire, dans les clades où l'activité est la plus forte, on observe de forts taux de variations, au niveau des réarrangements chromosomiques (par exemple, chez les mammifères et les squamates) qui se traduit par une évolution rapide de la structure du génome.

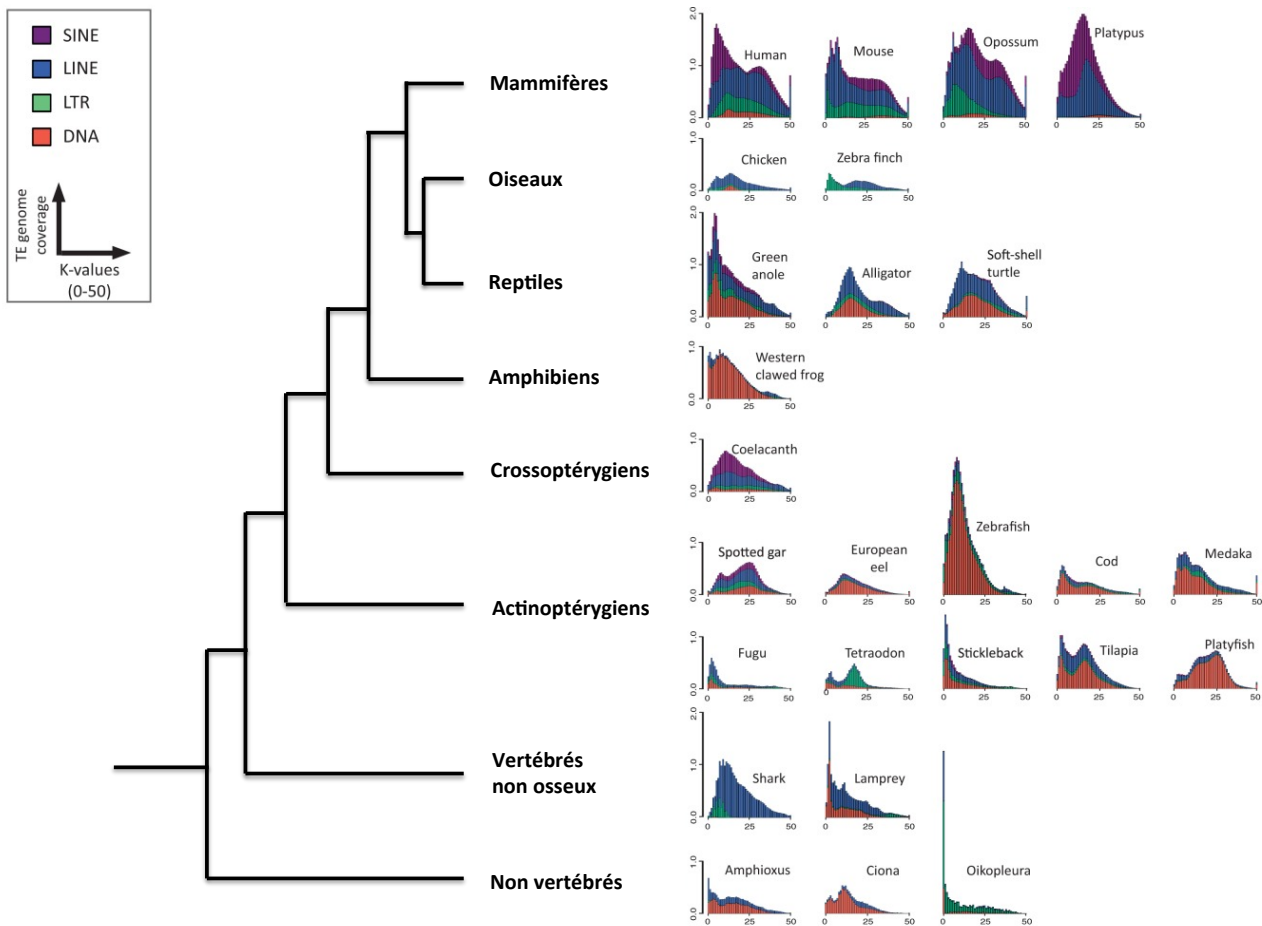


Figure 19. Représentation des éléments transposables selon la distance de Kimura dans les génomes de vertébrés et non vertébrés

La figure montre l'âge des grandes familles d'ET (distance de Kimura sur l'axe des abscisses). Les graphiques représentent la couverture du génome par chaque type d'ET (*ADN*, *SINE*, *LINE* et rétrotransposons à *LTR*) dans différents génomes. Les familles ont été regroupées selon la distance associée à leur séquence consensus (axe des abscisses). Les copies situées à gauche des graphiques divergent peu de la séquence consensus et correspondent à des copies récentes. Les séquences situées à droite des graphiques représentent des copies anciennes et dégénérées, hautement divergentes du consensus. Les paysages génomiques d'ET sont disponibles sur le site suivant: <http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html>.

III- Sexe et ET: l'art de façonner les génomes

1- Origine des chromosomes sexuels

L'émergence indépendante des chromosomes sexuels dans un grand nombre d'espèces eucaryotes semble être intimement liée à l'histoire des éléments transposables. La reproduction sexuée est le mode habituel de transmission de l'information génétique chez les eucaryotes, pour laquelle il existe une grande diversité de systèmes. Ces systèmes consistent en la fusion de gamètes à partir de deux individus au cours du processus de fertilisation, qui donne lieu à la formation d'un zygote. Les gamètes ancestraux étaient vraisemblablement petits et isogames (semblable en termes de morphologie). La transition vers un système anisogame (Maynard-Smith, 1958; Parker et al., 1972) est une étape cruciale au cours de l'évolution qui conduit à l'émergence de deux phénotypes sexuels distincts: la production de microgamètes mâles et de macrogamètes femelles (Figure 20). Le phénotype sexuel peut être déterminé par des causes environnementales, mais le plus souvent, il est déterminé génétiquement par un ou plusieurs locus localisés sur un chromosome autosomal ou sexuel. Les chromosomes sexuels sont le plus souvent hétéromorphes et sont apparus indépendamment dans plusieurs taxons. A ce titre, ils constituent une série intrigante de convergences évolutives. Parmi les couples de chromosomes sexuels les plus étudiés, il y a le système XY où les mâles sont hétérogames (XY) et les femelles homogames (XX); ainsi que le système ZW où les femelles sont hétérogames (ZW) et les mâles homogames (ZZ). Ces deux systèmes possèdent malgré tout un grand nombre de caractéristiques évolutives communes, notamment en termes d'ET, ce que nous allons présenter dans cette partie.

Les systèmes sexuels des espèces modèles comptent parmi les plus différenciés et ils donnent des perspectives intéressantes quant à leur évolution sur le long terme. Qu'il s'agisse des systèmes XY des mammifères et de la drosophile, ou des chromosomes ZW chez les oiseaux, il est aujourd'hui démontré qu'ils ont une origine autosomale (Figure 21) (Lahn and Page, 1999). On suppose qu'à l'origine, un locus déterminant le sexe, cesse de recombiner avec son homologue, soit à cause d'une inversion, soit que la sélection naturelle agisse en faveur de la cessation de la recombinaison à ce locus. Il s'agirait de l'événement initial à partir duquel, les chromosomes qui ne recombinent plus, commencent à dégénérer. Chez l'homme et le poulet de tels *loci* ont pu être identifiés, notamment par l'observation d'un nombre discret de clusters dont le degré de divergence est graduel, suggérant une série de réduction de la recombinaison dans ces régions (Lahn and Page, 1999). Chez l'humain, le degré de divergence le plus faible est observé au niveau de la région pseudo-autosomale (PAR) des chromosomes paralogues X et Y (Roos et al., 2004). Le degré de divergence augmente effectivement lorsque la distance à la région PAR augmente. Ce type de strate évolutive a été observé par ailleurs chez les plantes, les oiseaux et les champignons (Menkis et al., 2008; Nam and Ellegren, 2008; Nicolas et al., 2005).

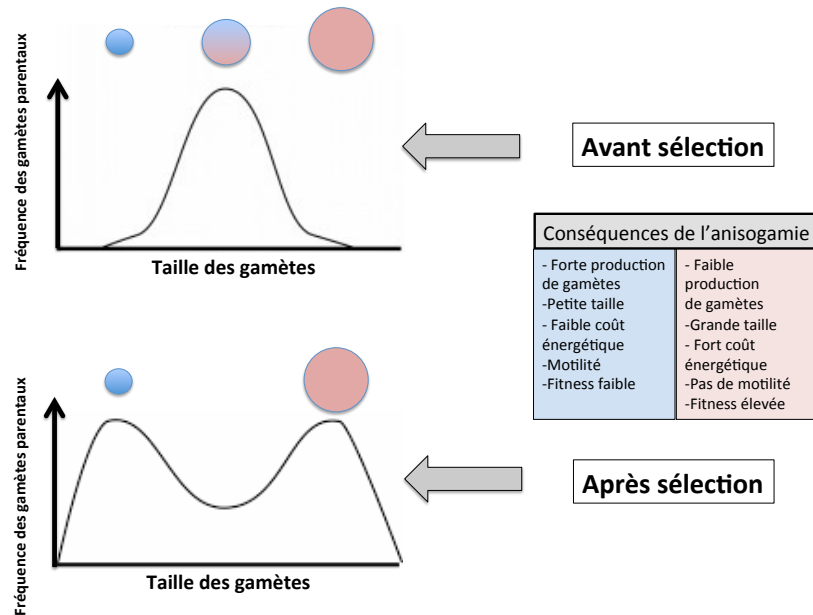


Figure 20. Anisogamie: l'évolution des genres mâle et femelles

La Figure illustre les forces sélectives auxquelles ont été soumis les gamètes ancestraux (petits et isogames). Il s'agit d'une sélection disruptive favorisant l'émergence de phénotypes extrêmes, ici au niveau de la taille des gamètes parentaux. Les gamètes femelles (rose) sont caractérisés par leur grande taille et de grandes réserves énergétiques de manière à optimiser les ressources disponibles pour la progéniture. Cependant, le coût investi pour obtenir de tels gamètes est supposé considérable et nécessite donc un compromis évolutif. Ce compromis se traduit *via* l'émergence de petits gamètes produits en très grand nombre, pour lesquelles les ressources sont très limitées mais qui possèdent une très grande taille de population, ce qui multiplie les chances de fertilisation. Au terme du conflit entre ces deux types de gamètes, il semble que les petits gamètes (très nombreux) vont entrer en compétition pour fertiliser un gros gamète mais assurant une fitness maximale pour la descendance (Trivers, 1972).

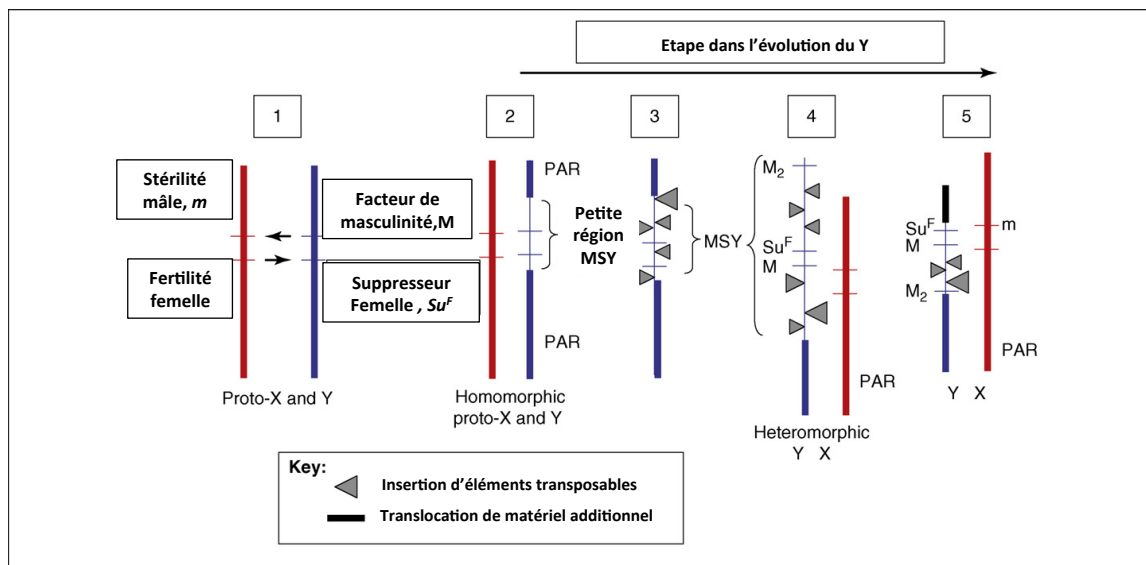


Figure 21. Schéma des étapes de l'évolution des chromosomes sexuels hétéromorphes

Les chromosomes sexuels sont formés à partir d'autosomes ordinaires qui contiennent un contenu en gène identique. La première étape dans l'évolution des chromosomes sexuels hétéromorphes est l'acquisition d'un locus déterminant du sexe sur un chromosome proto-sexe, comme un gène déterminant les mâles. Dans la première étape, une paire de chromosomes proto-sexe homomorphe contient une petite région contenant les gènes déterminants le sexe. Dans l'étape 2, la recombinaison est restreinte à cette région du chromosome Y (Région MSY, « male-specific region of the Y chromosome »). Les étapes 3 et 4 incluent l'expansion et les réarrangements de la région Y non-recombinante, causant de l'hétéromorphisme et l'addition de gènes spécifiques mâles. De la dégénérescence génétique a également lieu, avec la formation de pseudogènes sur le chromosome Y et l'évolution de la compensation de dosage. Les réarrangements (dus à la recombinaison entre des éléments transposables avec des séquences similaires localisées à différentes positions sur le chromosome) peuvent alors causer une perte majeure de matériel génétique, aboutissant à la réduction de la taille du Y et la perte de nombreux gènes. Adapté de (Charlesworth et al., 2005).

2- Conséquences évolutives de la différenciation des chromosomes sexuels

A - L'accumulation des ET chez les chromosomes sexuels anciens

En l'absence de recombinaison, les modèles prédisent une accumulation d'ET sur les chromosomes sexuels (Boutin et al., 2012; Dolgin and Charlesworth, 2006; Dolgin and Charlesworth, 2008). Une série grandissante d'exemples est en accord avec ces précédents modèles, où les chromosomes hétéromorphes ressemblent à de véritables « nids » à transposons. Dans une espèce avec des chromosomes sexuels différenciés de type XY, les effectifs efficaces de ces chromosomes sont égaux à 1/4 et 3/4 par rapport à ceux des autosomes, respectivement pour le Y et le X. On s'attend à ce que la sélection naturelle y soit moindre et donc à une fixation de mutations par dérive bien plus importante que sur les autosomes.

Dans les espèces modèles, la différenciation des chromosomes sexuels est telle que le chromosome hémigamétique ressemble souvent à un désert génétique principalement constitué de répétitions non codantes. Chez *D. melanogaster*, le chromosome Y est complètement dégénéré, très fortement hétérochromatique et constitué essentiellement d'ET et séquences microsatellites (Bachtrog, 2013). De même, chez les papillons du genre *Bombyx*, le chromosome hétérogamétique W accumule un grand nombre de rétrotransposons autonomes et de morceaux d'ET non autonomes (e.g. les 16,7 kb chez *Bombyx mandarina*) (Abe et al., 2010).

Chez les primates, le biais d'accumulation d'ET entre chromosomes autosomaux et sexuels a été mesuré. Ces études ont porté sur des lignées spécifiques des primates, les *Alu* (sous-famille de *SINE*) et les *LINE*, qui sont issus d'une vague d'amplification considérable il y a 40 millions d'années environ. Chez l'orang-outan, l'homme et le chimpanzé, on constate une forte densité d'*Alu* et de *LI* sur le chromosome X comparé aux autosomes (Kvikstad and Makova, 2010). Chez l'homme et le chimpanzé, espèces pour lesquels le chromosome Y a été séquencé, la densité d'*Alu* est élevée sur le chromosome Y, intermédiaire sur le chromosome X et faible sur les autosomes. Il n'existe chez l'homme, qu'une lignée d'ET (de type *LI*) pour laquelle la densité la plus forte est retrouvée sur le chromosome X. Ce cas particulier suggère que l'accumulation des ET par cliquet de Muller n'est vraisemblablement pas l'unique force évolutive qui contrôle la distribution des ET sur les chromosomes sexuels, ce que nous évoquerons dans la section suivante. Outre ces cas particuliers, on observe dans un large spectre d'espèces, ce type d'accumulation en éléments génétiques mobiles et autres séquences répétées. C'est le cas du chromosome Y chez la souris (Tucker et al., 2003), chez les *Felidae* (Stattery, 1998), chez les *Nototheniidae* (poissons de l'antarctique, (Ozouf-Costaz et al., 1996)), ou encore chez le saumon quinnat (Devlin et al., 1998). Cette tendance s'observe aussi chez les serpents (élapidae et colubridae) où l'on retrouve des chromosomes W très fortement hétérochromatiques, essentiellement constitués d'ADN répétés (Jones et al., 1987; O'Meally et al., 2010).

B - L'accumulation des ET chez les chromosomes sexuels émergents

D'autres taxons, comme les lézards et les poissons, sont apparus comme des modèles pour étudier le lien entre la nature répétée de l'ADN et la différenciation des chromosomes sexuels. Ils sont en effet caractérisés par une très grande diversité de systèmes, souvent jeunes, comparé aux autres vertébrés. La comparaison des chromosomes sexuels chez les espèces non modèles montre qu'en dépit de trajectoires évolutives drastiquement différentes, certains aspects fondamentaux subsistent en termes de dynamiques d'ET. Par exemple, les espèces *Leporinus reinhardti* et *Triportheus auritus* (Characiformes) présentent deux systèmes proches en temps de divergence avec cependant, des chromosomes sexuels dont la morphologie varie énormément. Chez ces espèces, la taille du W est extrêmement variable et ces variations s'expliquent essentiellement par l'accumulation massive de séquences répétées (Cioffi and Bertollo, 2012). Les espèces de lézards *Coleonyx elegans* et *Eremias velox*, possèdent respectivement un système XY euchromatique hétéromorphe et un système ZW hétérochromatique homomorphe (Pokorna et al., 2011). La paire de chromosomes euchromatiques hétéromorphe XY, ne montre aucun biais d'accumulation en séquences répétées et un contenu en ADN codant assez similaire, elle apparaît finalement peu différenciés en dépit de grandes variations de tailles (Pokorna et al., 2010). Cela contraste fortement avec la paire de chromosomes hétérochromatique homomorphes ZW, chez *E. velox*, qui possède un W caractérisé par une perte importante de séquences géniques, compensée par une forte accumulation d'ADN satellites et autres séquences répétées. Chez *E. erythrinus* (ordre des characiniiformes - poissons à dents), l'état des chromosomes sexuels est très variable entre populations de la même espèce. Certaines populations ne possèdent pas de chromosomes sexuels identifiables, tandis que d'autres, possèdent des systèmes multiples de type X_1X_2Y , apparus très récemment. L'activité des ET montre des différences remarquables entre populations d'une même espèce. Les copies de l'élément *Rex3* sont compartimentalisées dans les centromères de deux chromosomes uniquement pour les populations sans chromosomes sexuels différenciés. Dans l'autre cas, l'élément est distribué sur l'ensemble des 22 chromosomes et l'origine du Y dans ces populations semble liée à l'amplification de *Rex3*. En effet, le Y dérive d'une fusion de deux chromosomes acrocentriques nettement enrichis en *Rex3* au niveau de leur télomère. L'origine de la différenciation sexuelle des chromosomes que l'on observe entre les deux types de populations, est concomitante à l'explosion des éléments *Rex3*, qui semble précéder la dégénérescence physique et fonctionnelle des chromosomes hémigamétiques. Ces exemples illustrent la dynamique d'accumulation de séquences répétées, qui en somme, paraît assez bien corrélée au niveau de dégénérescence du chromosome hémigamétique.

C - Les ET présents dans la région du déterminisme sexuel (SD)

Chez les espèces dont le phénotype sexuel est déterminé de manière stable, le déterminisme du sexe est généralement présent génétiquement à un *locus* au moins. On appelle ce type de région SD pour « sex determination ». Au niveau de ces régions, on constate fréquemment des « hotspots » d'insertions d'ET, notamment chez des espèces pour lesquels le chromosome Y est encore jeune. C'est le cas chez le *medaka* et

Xiphophorus maculatus (Nanda et al., 2000; Volff et al., 2000; Volff et al., 1999). En effet, chez cette dernière espèce, une forte accumulation d'éléments de type *XIR* (LTR) et *Rex* (non LTR) a été détectée dans la région SD du chromosome Y alors qu'ils sont absents dans la région homologue du X. C'est aussi vrai pour la papaye (Liu et al., 2004), dont les chromosomes sexuels sont encore très peu différenciés, de même que chez l'espèce *Cironomoud Thummi* (diptère) qui montre une forte accumulation d'ET et de séquences répétées uniquement chez les mâles dans la région du SD (Kraemer and Schmidt, 1993). L'accumulation en ET intervient donc vraisemblablement très tôt, comme cela a été observé au locus du « mating-type » chez le champignon *Microbotryum violaceum* (Hood et al., 2004). Une fois que les chromosomes X et Y cessent de communiquer, il semble que la dégradation du Y soit rapide et vraisemblablement facilitée par une forte activité des ET, comme l'a montré expérimentalement Rice chez *D. melanogaster* (Rice et al., 2006).

3- Dégénérescence et dégradation

Outre l'accumulation d'ET, on note une véritable dégénérescence fonctionnelle au niveau des chromosomes hémigamétiques. Par exemple chez l'homme, le chromosome Y n'arbore que 78 gènes codant qui ne codent que pour 27 protéines différentes (Skaletsky et al., 2003), ce qui représente moins d'1 % de ce que l'on observe sur les autres chromosomes. Ceci contraste avec la forte densité en éléments génétiques mobiles présents sur le chromosome Y relativement aux autres chromosomes (Erlandsson et al., 2000). En effet, environ 30 % de la région euchromatique du chromosome Y correspond à des séquences dégénérées comportant des similitudes avec des séquences présentes sur le chromosome X. Environ 10 % de ces séquences sont issues de rétrotranspositions récentes (< 3Ma) et possèdent en moyenne deux fois plus d'ET que leurs séquences homologues trouvés sur le X. Ce « syndrome de dégénérescence » du chromosome Y est observé chez un grand nombre d'espèces non apparentées phylogénétiquement et semble être une conséquence directe de l'absence de recombinaison sur le chromosome Y.

Au niveau de la paire de chromosomes X/Y chez *D. melanogaster*, par exemple, il n'existe plus de séquences homologues reconnaissables entre les deux chromosomes. Le Y ne contient plus qu'une dizaine de gènes, constitués d'énormes introns pour la plupart d'entre eux (Carvalho et al., 2001). De manière comparable au chromosome Y humain, la plupart des gènes liés au chromosome Y ont été recrutés à partir des chromosomes autosomaux et possèdent des fonctions spécifiques du mâle (Carvalho, 2002). Chez l'espèce *D. pseudobscura*, pour laquelle l'origine du chromosome Y est plus récente (10 millions d'années), les gènes originaires du Y ont rapidement été transloqués sur les autosomes (Figure 22). L'espèce *D. miranda* possède un chromosome Y apparu plus récemment encore (1 million d'années) pour lequel on peut suivre l'état de dégradation. On y observe déjà l'accumulation de plusieurs classes de rétrotransposons sans LTR de type *worf*, *spock*, *TRIM* et *TRAM*. De plus, plusieurs de ces copies sont spécifiquement insérées dans des gènes tels que *Lcp* et *exul*, de manière à fortement diminuer leur expression transcriptionnelle (Bachtrog, 2003; Steinemann and Steinemann,

1998). La perte de fonction et la dégénérescence du chromosome hémigamétique caractérisent les premières étapes de la différenciation de la paire de chromosomes sexuels, au cours desquelles les ET semblent activement impliqués (Figure 22).

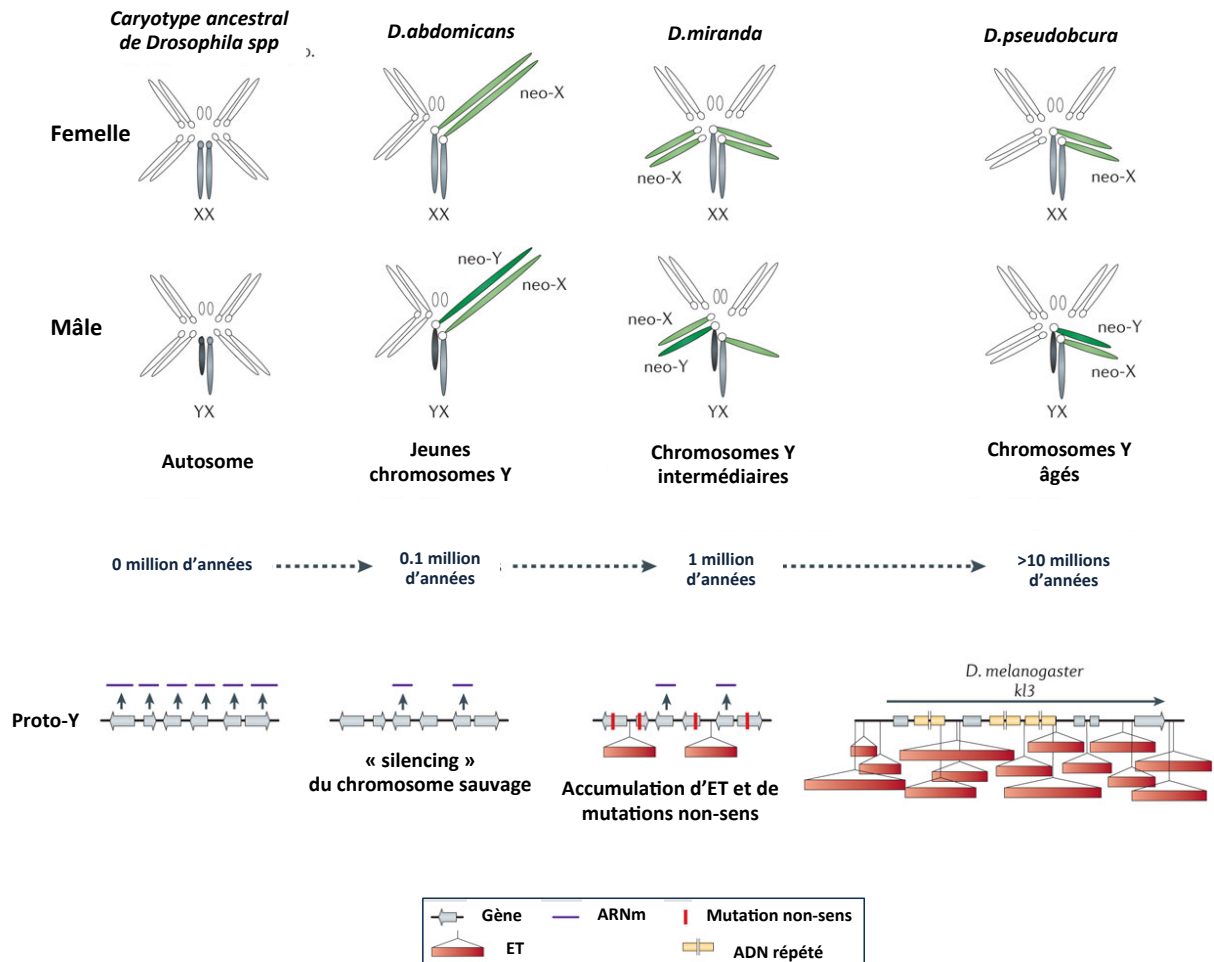


Figure 22. Les chromosomes néo-sexes chez la drosophile

Les chromosomes néo-sexes (en vert) sont formés par fusions des autosomes avec les chromosomes sexuels ancestraux (en gris). Le néo-X et néo-Y portent un ensemble de gènes identiques au départ (0 million d'année). A ce jour, trois espèces de drosophiles avec des néo-Y de différents âges ont été étudiées en détails et donnent une image des changements moléculaires associés à la différenciation du Y. Chez le néo-Y de *D. abdominalis*, on observe une diminution de l'expression de gènes sauvages probablement causée par des modifications épigénétiques résultant d'un changement de structure de la chromatine (0.1 million d'années). Une altération générale de l'architecture du génome du néo-Y a eu lieu il y a seulement 1 million d'années chez *D. miranda*. En effet, on observe chez le néo-Y de cette espèce, de nombreuses mutations telles que des codons stops, des décalages du cadre de lecture, de larges délétions et une accumulation de répétitions d'éléments transposables. La perte de fonction de gènes est accompagnée par le début de l'hétérochromatinisation du néo-Y. Après 15 millions d'années, on ne retrouve presque plus de similitudes entre le néo-Y et le chromosome dont il est issu. Ainsi, le néo-Y devient entièrement hétérochromatique (par exemple chez *D. pseudoobscura*). Adapté de (Bachtrog, 2013).

Les arguments théoriques concernant la réduction de taille du chromosome Y par rapport aux autres chromosomes (chez *D. melanogaster* par exemple) supposent un effet de la sélection naturelle contre le taux de recombinaison ectopique lié à l'accumulation d'ET dupliqués (Charlesworth et al., 1994). Ce modèle d'échange par recombinaison ectopique est soutenu expérimentalement par Petrov (Petrov et al., 2011) qui montre chez *D. melanogaster*, que les éléments de longue taille sont rarement fixés dans les populations naturelles, par contraste avec la persistance d'éléments de petites tailles plus fréquemment fixés. Plus précisément, il montre que les forces de sélection purifiantes sont d'autant plus fortes que le taux de recombinaison, la taille et le nombre de copies des éléments augmentent. Une autre hypothèse concernant la réduction de taille du chromosome Y est basée sur l'observation, chez l'homme, que les ET peuvent causer directement la perte de certains gènes. Au regard du chromosome 4 chez *D. melanogaster*, ponctiforme et très riche en ET, il semble que la dégénérescence physique et fonctionnelle des chromosomes sexuels soit accélérée par l'accumulation des ET (Vicoso and Bachtrog, 2013).

4- Les ET dans la lignée germinale mâle

A - La surexpression de l'activité des ET dans les testicules

Il a été suggéré depuis longtemps que les ET montraient une tendance à la suractivation dans les testicules des mammifères (Weiner et al., 1986), ce qui corrèle assez bien avec la fixation d'ET à fortes fréquences sur le chromosome Y. Le séquençage du génome humain confirme cette tendance au regard de la distribution des ET sur les chromosomes (Jurka et al., 2004; Jurka et al., 2002). L'exemple le mieux décrit concerne une sous-famille d'*Alu* (*AluY*) dans la lignée humaine qui montre une très forte accumulation sur le chromosome Y (deux fois supérieure à la densité observée sur les autosomes), datée d'environ 6 millions d'années. Cette forte accumulation sur le chromosome Y suggère que les ET étaient préférentiellement actifs dans la lignée germinale mâle. Par ailleurs, la distribution de cette famille sur les autres chromosomes est proportionnelle au temps passé dans la lignée germinale mâle, c'est-à-dire que la densité des *AluY* sur le chromosome X est de 2/3 celle observée sur les autosomes. Cette tendance est retrouvée pour d'autres rétrotransposons, dans d'autres espèces de mammifères. C'est le cas des *SINE*, *LINE* et ERV (rétrovirus endogènes) qui montrent le même type de distribution sur l'X de la souris (Nellaker et al., 2012).

La surexpression des ORF de l'élément *LINE-1* dans les testicules de mammifères est démontrée depuis longtemps (Branciforte and Martin, 1994; Sassaman et al., 1997; Singer et al., 1993). De même, chez *D. melanogaster*, plusieurs exemples décrivent une surexpression des ET dans la lignée germinale mâle. C'est le cas des rétrotransposons GATE (Kogan et al., 2003), *copia* (Pasyukova et al., 1997), *1731* (Haoudi et al., 1997), *micropia* (Lankenau et al., 1994). Un autre cas particulier est le rétrovirus endogène *OvexI*, présent

depuis au moins 100 millions d'années chez les oiseaux, qui est surexprimé fortement dans les gonades mâles au moment de la différenciation morphologique.

Le transcriptome d'une espèce non modèle, la pétoncle, montre également une surexpression globale d'une grande diversité d'ET dans les tissus associés à la spermatogenèse (Llera-Herrera et al., 2013). De même, l'étude transcriptomique du coelacanth révèle une forte surexpression de *LINE* (*CRI*), *L2* et *SINE* spécifiquement dans les testicules. Cette surexpression corrobore les vagues d'amplifications récemment observées chez plusieurs espèces du genre *Latimeria* (Forconi et al., 2014).

Ces exemples dans diverses espèces non apparentées phylogénétiquement, amènent à penser que cette surexpression est un phénomène global en lien avec le processus de spermatogenèse. Plus globalement, on peut spéculer que les testicules seraient une niche environnementale favorable à l'activité des ET. Lorsque l'on regarde les exemples d'adaptations acquises au cours de la différenciation des chromosomes sexuels, il semble en effet, que l'activité des ET dans la lignée germinale mâle occupe souvent un rôle déterminant.

B - La surexpression dans les testicules est favorisée par la sélection sexuelle

Le séquençage des 12 génomes de drosophiles a montré que les gènes présents sur le chromosome Y avaient une origine autosomale et ont été acquis il y a moins de 63 millions d'années, après la divergence des sous-genres *Sophophora* et *Drosophila* (Koerich et al., 2008). De manière surprenante, l'ensemble de ces gènes semble avoir été sélectionné pour des fonctions spécifiques chez les mâles. En effet, la plupart d'entre eux ont des effets forts sur le déterminisme du sexe ou sur la stérilité mâle et sont surexprimés chez les mâles. Une étude comparative des chromosomes Y de *D. melanogaster* et *D. pseudoobscura* a montré qu'aucun des gènes présents n'étaient similaires (Carvalho and Clark, 2005). Dans les deux espèces, ces gènes ont été dupliqués par rétrotransposition à partir d'un ARNm autosomal ou lié au chromosome X. Il a été constaté par ailleurs, qu'indépendamment de leur origine chromosomique, les gènes rétrotransposés ont presque toujours une expression testicule-spécifique (Meisel et al., 2009).

Les chromosomes Z du poulet et X de l'homme ont évolué indépendamment, mais présentent des caractéristiques communes, acquises de manière convergente au cours de la différenciation des chromosomes sexuels. Les deux chromosomes ont subi d'importants remodelages: la perte de dizaines de gènes accompagnée d'une forte accumulation de séquences intergéniques et de *LINE*. Malgré leur faible densité en gènes, chaque lignée chromosomique a acquis des centaines de gènes en multicopies, surexprimés pour la grande majorité dans les testicules (Bellott et al., 2010).

L'effet de la sélection sexuelle est aussi suggérée *via* le recrutement de séquences répétées ou dérivées d'ET, qui présentent des profils d'expressions particulièrement élevés dans les testicules et/ou une surreprésentation sur le chromosome hémigamétique. C'est le cas, par exemple, de la famille de transposons à ADN, *Zipuston*. Cette famille code une protéine qui possède un domaine d'interaction à l'ADN ainsi qu'un

domaine SUMO qui intervient dans la modification post-traductionnelle des protéines. L'expression de cette transposase est conservée dans la plupart des espèces du genre *Xiphophorus* et présente une forte expression dans les testicules. Chez l'espèce *X. maculatus*, son expression dans les testicules est très élevée et probablement liée à la présence de plusieurs copies sur le chromosome Y, là où les protéines surexprimées dans les testicules sont souvent localisées (Skaletsky et al., 2003). Un autre exemple d'amplification de familles géniques sur le chromosome Y a été observé chez *X. maculatus*. Il s'agit d'une famille d'hydrolase (SGNH), dont l'expression est très forte dans les testicules. D'une part, ce gène est apparenté aux *LINE* de type CR1, bien qu'il ne code pas pour une reverse-transcriptase fonctionnelle. D'autre part, chacune des duplications est flanquée par un ET de la famille *Hélitron*. Ce gène a vraisemblablement été « capturé » par l'hélitron et dupliqué successivement sur le chromosome Y par transposition (mécanisme de *rolling-circle*). Dans cet exemple, un transposon aurait directement favorisé l'émergence d'une nouvelle famille de protéines essentielle au processus de spermatogenèse, caractérisée par sa forte expression dans les testicules.

Par ailleurs, la sélection sexuelle semble avoir favorisée le recrutement de motifs répétés sur les chromosomes sexuels, qui agissent comme des plateformes d'interactions pour des protéines se liant à l'ADN. Par exemple, chez *Schizosaccharomyces pombe*, un facteur homologue de SRY chez l'homme, interagit à plusieurs niveaux avec des ET pour déterminer le sexe mâle (Matsuda et al., 2011). Tout d'abord, la reconnaissance des *loci* mâles est facilitée par un enrichissement local en LTR, ensuite, l'activation des *loci* impliqués dans la détermination du sexe, requiert l'intervention d'une transposase domestiquée (*via* son domaine d'interaction à l'ADN dérivé d'un transposon *pogo*, connu pour faciliter la formation des centromères). Un autre exemple est celui de la séquence répétée *banded krait mircosatellite* (*Bkm*), retrouvée sur les chromosomes hémigamétiques des plantes et des animaux, homme inclus (Jones and Singh, 1981). La synténie de cette répétition en tandem n'est pas conservée, suggérant qu'elle a été recrutée sur le chromosome hémigamétique à de multiples reprises au cours de la différenciation des chromosomes sexuels. Il a été montré en outre, que des protéines impliquées dans la gestion de l'hétérochromatine pouvaient se lier à cette séquence *Bkm* et que ces dernières se trouvaient systématiquement surexprimées dans les gonades, chez le serpent, le ver à soie, la souris et l'homme (Priyadarshini et al., 2003; Singh and Jones, 1986; Singh et al., 1994).

Dans ces exemples, il semble que la sélection sexuelle façonne de manière convergente l'évolution du chromosome requis pour la différenciation mâle. On observe systématiquement une fuite des gènes vers les autosomes puis le recrutement tardif de gènes pro-mâles. Au regard des événements de rétrotranspositions, de l'accumulation des séquences *cis*-régulatrices ou du recrutement de séquences interagissant avec l'hétérochromatine, les séquences répétées semblent, dans leur ensemble, créer un contexte favorable à l'émergence de fonctions mâle-spécifiques.

5- Les adaptations liées au sexe médiées par les ET

Plusieurs exemples suggèrent fortement que l'accumulation des ET a été favorisée et participe directement à l'acquisition de nouvelles fonctions, liées au sexe notamment. Parmi ces exemples, certains montrent clairement que le déterminisme génétique du sexe est directement lié à une insertion d'ET. Le gène *Dmrt1* du medaka (*Oryzias latipes*) est conservé chez les protostomiens chez lesquels il est responsable du déterminisme sexuel (homologue du gène *double-sex* des drosophiles et de *mab-3* chez *C. Elegans*). Une copie dupliquée de ce gène est retrouvée sur le chromosome Y du medaka et il a été montré qu'il était le gène essentiel dans la régulation du développement sexuel. Cette duplication est caractérisée par une insertion d'ET dans sa région promotrice: il s'agit de l'élément *Izagani* qui dérive de l'élément *P*. Cette insertion est absente dans la région homologue du chromosome X et permet l'expression de *Dmrt1* spécifiquement dans les testicules pour la différenciation en sexe mâle (Herpin et al., 2009). Un autre cas d'ET recruté pour induire le déterminisme sexuel a été décrit chez le melon. L'insertion d'un transposon à ADN (de la famille *hAT*) dans la région promotrice du facteur de transcription *CmWIP1*, permet son expression sexe-spécifique. Dans cet exemple, le déterminisme est de nature épigénétique, puisque l'insertion de l'élément entraîne la propagation et le maintien de la méthylation de l'ADN dans cette région de manière à initier localement le processus de différenciation sexuelle (Martin et al., 2009).

L'espèce *Bombyx mori* possède un chromosome W dont le rôle est déterminant pour la détermination du sexe femelle, suggérant la présence d'un gène féminisant sur ce dernier. Cependant, le chromosome W est constitué essentiellement d'ET et aucun gène codant n'a pu être identifié à ce jour. Kiuchi et son équipe ont finalement montré qu'un précurseur (*fem*) de *piARN* initiait la voie de signalisation femelle dans l'embryon (Kiuchi et al., 2014). Il a été montré d'une part, que l'inhibition de ce précurseur conduit inévitablement à la différenciation vers le sexe mâle et d'autre part, que ce dernier permet la production de transcrits alternatifs du gène *double-sex* (*Bmdsx*) intervenant spécifiquement dans la cascade de différenciation sexuelle.

Sur le chromosome X de l'homme, on trouve une forte densité d'élément *LI* (supérieure à celle du chromosome Y) qui suggère qu'ils ont été fixés de manière non aléatoire par la sélection naturelle. On retrouve préférentiellement les *LI* à proximité des gènes et à plus forte fréquence près des gènes éteints au cours du processus d'inactivation du X (Abrusan et al., 2008; Carrel et al., 2006). Il a été suggéré que les séquences *cis*-régulatrices présentes dans l'élément stabiliseraient le recrutement des protéines impliquées dans le processus d'inactivation. Ce recrutement spécifique à la lignée humaine, reste néanmoins le seul cas décrit chez les vertébrés. En effet, chez le poulet, de telles séquences ont été recherchées en vain. On observe bien une accumulation d'éléments de type *CRI* (*LINE*), mais ceux-ci ne s'accumulent pas à proximité des gènes soumis à la compensation de dose (Ellegren et al., 2007; Melamed and Arnold, 2007). De la même façon, chez l'ornithorynque, il n'y a pas d'accumulation de *LINE* sur le chromosome X (Deakin et al., 2008). Néanmoins, les effets de dosage chez ces espèces sont nettement moins importants que chez l'homme.

Chez *D. melanogaster*, l'enrichissement en ET sur le chromosome X, relativement aux autosomes n'est pas observé (Kaminker et al., 2002), ce qui est probablement lié aux effets délétères associés à la recombinaison ectopique. Cependant, chez *D. miranda* (environ 30 millions d'années de divergence avec *D. melanogaster*), des ET ont été recrutés sur le chromosome X et jouent un rôle essentiel dans la compensation de dose (Figure 23). En effet, des dizaines de sites de liaisons au complexe de compensation de dose (MSL) ont été acquis par des insertions répétées d'hélitrons tout le long du chromosome. De manière intéressante, on n'observe plus de nouvelles insertions sur ce chromosome et les copies présentes ont acquises des mutations ponctuelles qui augmentent significativement l'affinité pour le complexe MSL.

Cette liste, non exhaustive d'exemples, illustre assez bien la relation étroite qu'il existe entre le maintien des ET dans le génome et leurs effets sur l'évolution des chromosomes sexuels. Il semble que les ET participent activement, bien que de manière stochastique, à l'acquisition de nouvelles fonctions au cours de la divergence des chromosomes sexuels.

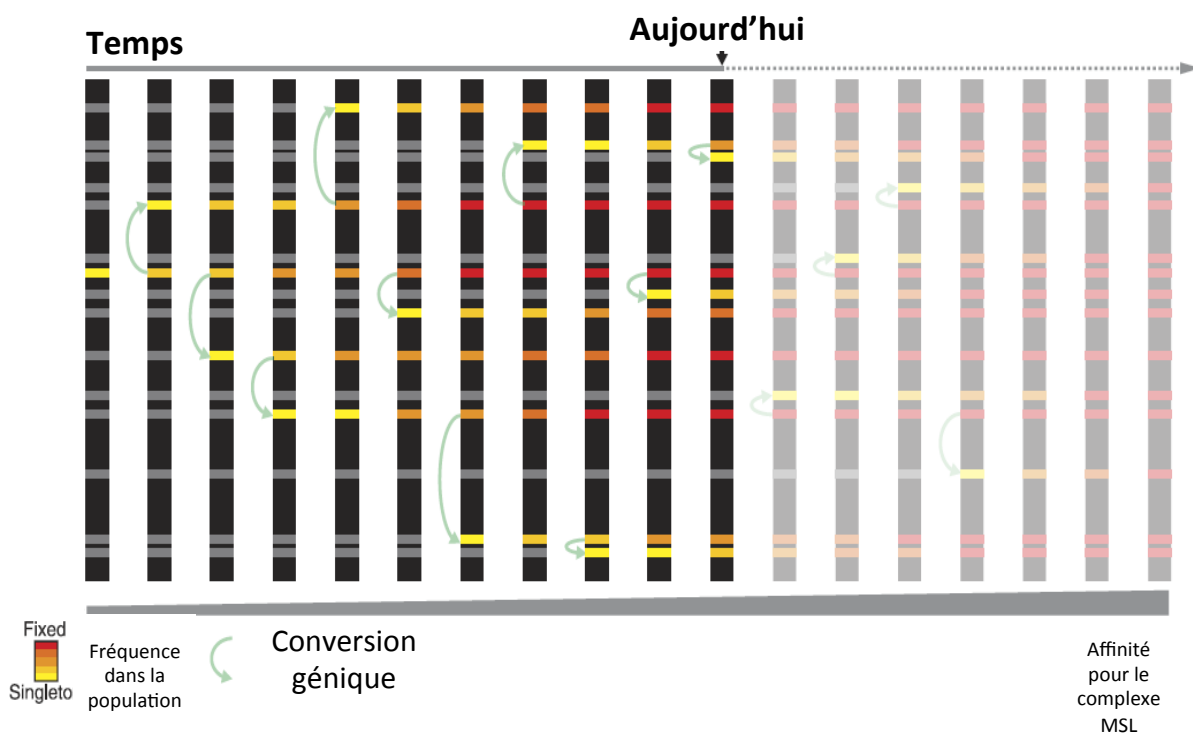


Figure 23. Exemple d'adaptation liée à l'accumulation d'un ET sur le chromosome Y chez *D. miranda*

La figure schématise le scénario évolutif ayant favorisé l'émergence du processus de compensation de dosage chez les mâles, suite à l'invasion massive d'un hélitron sur le chromosome Y. Les rectangles noirs représentent l'état du chromosome Y au cours du temps. Les lignes grises correspondent à des insertions d'hélitrons caractérisées par la présence de régions riches en Thymine. Ces régions riches en T, créent un contexte favorisant l'affinité du complexe MSL (Male Specific Lethal – permettant la compensation de dosage). Au temps initial, une mutation par substitution apparaît dans ces régions et augmente l'affinité du site pour le complexe de compensation de dosage (ligne jaune). Au cours des générations, cette mutation avantageuse augmente en fréquence dans la population (gradient du jaune au rouge) jusqu'à sa complète fixation (rouge). En parallèle, on observe un phénomène de conversion génique (flèche verte) qui remplace l'haplotype standard (gris) par le nouvel haplotype (jaune). Par ce phénomène, on observe l'accumulation, dans la population, de sites ayant une forte affinité pour le complexe MSL et la région acquiert progressivement la capacité à effectuer la compensation de dosage. Adapté de (Vicoso and Bachtrog, 2013).

IV- Régulation des ET dans la lignée germinale

1- Le maintien de l'intégrité du génome: les *piARN* dans la lignée germinale

La découverte du concept moléculaire d'ARN interférents (Fire et al., 1998) a eu des conséquences fondamentales sur la compréhension des moyens mis en place par le génome pour stabiliser sa composante en éléments génétiques mobiles et/ou de se protéger des séquences exogènes telles que les virus.

On distingue plusieurs classes d'ARN interférents: les *microARN*, les *siARN* et les *piARN*. Les protéines Argonautes tiennent un rôle essentiel au cours du processus de production des ARN interférents. Les membres de cette famille sont trouvés dans tous les domaines du vivant et sont impliqués dans une grande diversité de voies cellulaires allant de l'inhibition transcriptionnelle des ARNm dans le cytoplasme, à la compaction hétérochromatinienne d'une multitude de loci dans le noyau. Phylogénétiquement, les protéines Argonaute eucaryotes se regroupent en trois clades: le clade AGO, PIWI et WAGO (spécifique des nématodes). Le clade AGO interagit spécifiquement avec les *microARN* et les *siARN* qui sont dédiés à la répression post-transcriptionnelle des gènes (Figure 24). Les membres du clade PIWI interagissent principalement avec des petits ARN un peu plus longs, les *piARN*.

Un grand nombre de génomes procaryotes contient des homologues d'Argonaute pour lesquels, les domaines *MID* et *PIWI* sont conservés (Figure 25). A la différence de leurs homologues eucaryotes, elles ont la caractéristique de s'associer préférentiellement à un brin guide ADN plutôt qu'un ARN. Cependant, ces homologues semblent aussi impliqués dans la défense du génome contre les éléments génétiques mobiles (Marakova et al, 2009). Au niveau structural, ces protéines sont caractérisées par la conservation de quatre domaines protéiques. Le domaine *MID* forme une poche permettant la liaison d'une courte séquence d'acide nucléique et interagit spécifiquement avec l'extrémité 5'-Phosphate d'un ARN. Ce domaine est indispensable à la reconnaissance du petit ARN, à sa stabilisation sur la protéine et sa bonne conformation pour un clivage effectif. Le domaine *PIWI* possède une activité Rnase-H qui correspond à l'activité de coupure des protéines AGO. Le domaine *PAZ* permet la protection du petit ARN de la dégradation *via* son interaction avec l'extrémité 3'. Le domaine *N* est quant à lui essentiel à la reconnaissance du site de coupure et à la dissociation du duplex ARN chargé, c'est-à-dire qu'il permet au brin possédant l'extrémité 5' la moins stable d'être retenue comme brin guide (Swarts et al., 2014a).

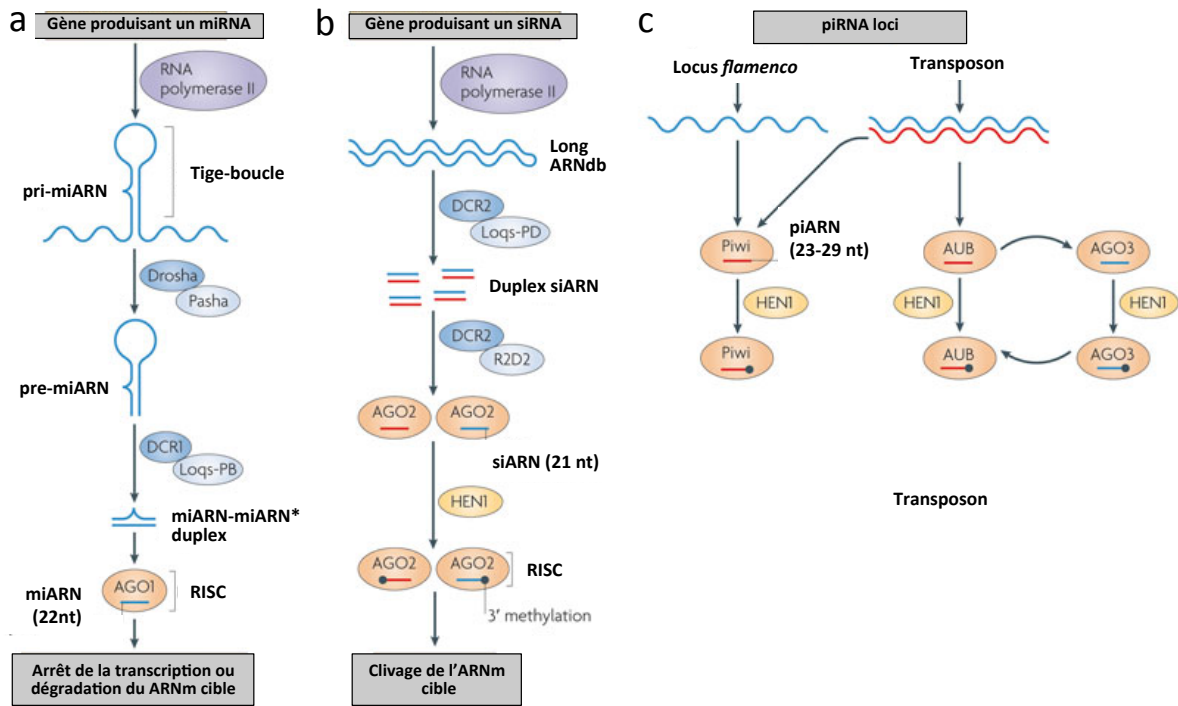


Figure 24. Schéma représentant les mécanismes de synthèse des petits ARN interférents

a) Le pri-*miARN* est transcrit par l'ARN Polymérase II à partir d'un gène codant un ARNm. Dans le noyau, le pri-*miARN* est ensuite stabilisé par la protéine Pasha (chez les mammifères) et processé par une ribonucléase de type III appelée Droscha pour former un pré-*miARN*. Ce dernier est ensuite exporté dans le cytoplasme où il est de nouveau processé par Dicer1 (Dcr1) associé à la protéine Loquacious-isoform PB (Loqs-PB, chez les mammifères), aboutissant à la formation d'un duplex *miARN*. Le *miARN* simple brin (22 nt) est ensuite associé à Ago1 pour former le complexe RISC (*RNA-Induced Silencing Complex*) qui sera responsable de l'arrêt de la transcription de l'ARNm cible ou de la dégradation de l'ARNm cible. b) Les *siARN* sont générés à partir d'un gène transcrit par l'ARN Polymérase II, premièrement, sous la forme d'un long ARN double brin. Ce dernier est processé par la RNase de type III, Dicer 2 (Dcr2), probablement associé à Loqs-PD, en petits duplex *siARN*. Ces duplex sont alors pris en charge de nouveau par Dcr2 associé à R2D2 qui charge les protéines Ago2 d'un brin de *siARN* formant alors un complexe RISC. c) Les *piARN* sont produits à partir d'un locus *flamenco* ou à partir d'un transposon. Contrairement aux autres voies de biogenèse, d'ARN interférents, celle des *piRNA* ne nécessite pas de RNase de type III. Les *piARN* produits sont principalement antisens et interagissent avec Piwi ou Aubergine (Aub). Les *piARN* sens interagissent avec Ago3 et collaborent avec Aub pour amplifier les *piARN*. La méthyle transférase Hen1, ajoute une modification 2'O methyl à l'extrémité 3' des *siARN* et des *piARN* après leurs interactions avec les protéines Ago. Issu de (Ding, 2010).

Protein	Guide	Target	Present in	Domain architecture
Short pAgo*	?	(DNA?)	Prokaryotes	Nuclease APAZ MID PIWI*
Long pAgo	DNA/(RNA)	DNA/RNA	Prokaryotes	N L1 PAZ L2 MID PIWI
Long pAgo*	RNA/(DNA)	DNA/(RNA)	Prokaryotes	Nuclease N L1 PAZ L2 MID PIWI*
eAgo	RNA	RNA	Eukaryotes	N L1 PAZ L2 MID PIWI
eAgo*	RNA	RNA	Eukaryotes	N L1 PAZ L2 MID PIWI*
PIWI-RE	(RNA)	(DNA)	Bacteria	Domain X MID PIWI
PIWI-RE*	(RNA)	(DNA)	Bacteria	REase DExD/H Domain X MID PIWI*
Med13	(RNA)	?	Eukaryotes	Med13-N MID PIWI*

Figure 25. Architecture des domaines des protéines de la superfamille des PIWI

La figure montre le haut niveau de conservation des différents domaines protéiques associés aux protéines du clade AGO. eAgo désigne les protéines Argonautes eucaryotes et pAgo leur homologues procaryotes. Bien que ces protéines peuvent s'associer à différents substrats (double d'ADN, d'ARN ou les deux), la majorité des domaines sont très conservés. Les domaines *MID* et *PIWI* sont retrouvés systématiquement malgré plusieurs milliards d'années de divergence. Issu de (Swarts et al., 2014b).

Chez les espèces animales sexuées, les protéines PIWI et les *piARN* s'expriment essentiellement dans la lignée germinale et sont majoritairement impliqués dans la répression des ET. Pour assurer la répression des ET, les *piARN* s'associent avec les protéines PIWI de manière à former un complexe (*piRISC*) qui peut reconnaître et inhiber les ARN cibles complémentaires (Malone and Hannon, 2009). Il a été montré par un grand nombre d'études génétiques, que les *piARN* et les protéines associées à leur production jouaient un rôle essentiel dans le développement et le maintien de la lignée germinale mâle et femelle.

Chez *D. melanogaster*, il existe trois membres de la famille PIWI : Piwi, Aubergine (Aub), Argonaute3 (Ago3). Une mutation dans l'un de ces gènes conduit, d'une part à une dérégulation globale des ET dans l'ovaire et d'autre part, au développement défectueux des gonades. Chez la souris, on compte aussi trois membres de la famille PIWI (Mili, Miwi et Miwi2). Une déficience au niveau des protéines Mili et Miwi2, conduit à l'activation des *LINE* et des rétrotransposons à LTR dans les testicules. Comme chez la drosophile, ces mutations sont associées à un arrêt de la gamétogenèse et une stérilité mâle complète. Malgré leurs homologues, ces protéines ont divergé et leur fonction n'est pas redondante dans la cellule. Par exemple, l'autre membre de cette famille (Miwi), n'induit ni la stérilité ni la dérégulation des ET. En outre, il est intéressant de constater que d'un sexe à l'autre, ces protéines n'occupent pas non plus la même fonction. L'étude en contexte mutant de Mili, Miwi, Miwi2, ne montre aucun effet majeur sur le développement de la lignée germinale femelle. Par analogie, il a été observé que les mutants Ago3 n'affectent que partiellement la stérilité mâle chez *D. melanogaster*, suggérant que ces protéines ont été recrutées de manière préférentielle dans l'un des deux sexes.

2- Les mécanismes de biogenèse des *piARN*

L'étude intensive des *piARN* pris en charge par les protéines PIWI chez la souris et la drosophile a permis de mettre en évidence leur origine génomique. Suivant leur origine génomique, les *piARN* ne suivent pas le même mode de production. Certains sont issus d'un processus de biogenèse primaire (Figure 26a) tandis que d'autres sont issus d'une boucle d'auto-amplification (la boucle ping-pong) (Figure 26b). Il semble que le maintien de ces deux processus soit indispensable pour une répression efficace des ET. Selon le modèle dominant, la biogenèse primaire fournirait le matériel initial pour cibler de multiples séquences génomiques apparentées aux ET. Intervient ensuite, la boucle d'amplification pour façonner le répertoire initial, en amplifiant la production de *piARN* ciblant les familles d'ET les plus actives. Ces deux processus de production semblent conservés chez l'ensemble des vertébrés (l'homme, le rat, le moustique, le poisson zèbre, la grenouille) mais aussi chez des espèces plus éloignées comme les diptères et les lépidoptères (Houwing et al., 2008; Kawaoka et al., 2009; Lau et al., 2006; Robine et al., 2009).

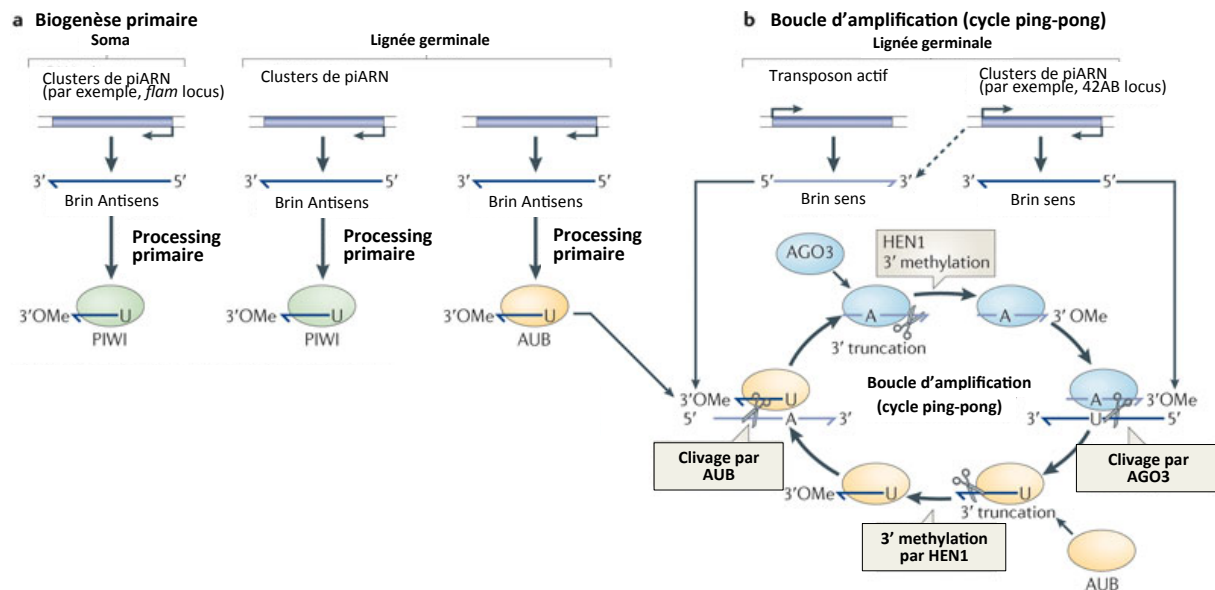


Figure 26. Mécanisme de biogénèse de *piARN* et leur amplification par la boucle ping-pong

a) Le transcrit antisens, issu du transposon et/ou des clusters de *piARN*, est traité en *piARN* par un mécanisme inconnu et est chargé sur Aubergine (Aub) ou Piwi. Les *piARN* dérivés du locus *flam* sont chargés exclusivement sur Piwi car *flam* n'est actif que dans le soma ovarien dans lequel, seul Piwi est exprimé. Les complexes *piRISC* (*piRNA-Induced Silencing Complexes*) produits via ce mécanisme, agissent comme un déclencheur de la boucle d'amplification ping-pong. b) Le mécanisme d'amplification (cycle ping-pong) est principalement impliqué dans l'activité Slicer de Aub et Ago3, mais pas de Piwi. Aub, associé avec le *piRNA* antisens, clive le *piARN* précurseur sur le brin sens. Cette réaction détermine et forme l'extrémité 5' des *piARN* qui sont chargés sur Ago3. Ago3, associé au *piARN* sens, clive le *piARN* précurseur antisens, générant l'extrémité 5' du *piARN* antisens qui est alors chargé sur Aub. L'extrémité 3' du *piARN* est formée par une nucléase inconnue puis, subit une 2'O méthylation par Hen1. Adapté de (Siomi et al., 2011).

A - La biogenèse primaire

Contrairement aux *microARNs*, pour lesquels on distingue quelques centaines de séquences individuelles, les premiers séquençages de *piARN* ont montré leur extraordinaire complexité. Il s'agit de plusieurs centaines de milliers, voire des millions de séquences individuellement distinctes. Par ailleurs, ces séquences ne présentent pas de caractéristiques communes, comme c'est le cas pour les *microARN*, à l'exception d'un fort biais d'uridine en position 1 de la séquence. De manière plus surprenante encore, lorsque l'on aligne cette grande diversité de séquences sur les génomes, on constate qu'elles ne proviennent que d'une très faible fraction de ce dernier. Chez *D. melanogaster*, les *piARN* s'alignent sur environ 3 % du génome et sont regroupés dans un nombre discret de *loci*. On appelle ces *loci* des « clusters » de *piARN*, qui contiennent de multiples séquences pouvant générer des *piARN* et ont une taille de quelques kilobases en général (jusqu'à 200 kb) (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006).

Au niveau structural, ces clusters ne présentent aucune structure secondaire particulière, contrairement aux clusters de *microARN* organisés en tige-boucle. On distingue cependant différents types de clusters en fonction de l'orientation des *piARN* qui s'y trouvent. Certains clusters s'alignent avec des *piARN* sur les deux

brins d'ADN (sens et antisens) suggérant que, dans ce type de régions, la transcription est bidirectionnelle. C'est le cas par exemple, du cluster 42AB chez *D. melanogaster* (Brennecke et al., 2007). Pour d'autres (e.g. le locus *flamenco* chez *D. melanogaster*), les *piARN* ne sont produits qu'à partir d'un seul des deux brins (le brin antisens par rapport à l'ARNm de l'ET, le plus souvent). Il a été montré par ailleurs, que de longs ARN précurseurs étaient transcrits au niveau des clusters de *piARN*, qui par la suite sont processés en *piARN* matures d'une taille entre 23 et 29 nucléotides (Aravin et al., 2006; Aravin et al., 2007; Brennecke et al., 2007; Girard et al., 2006). Pour le locus *flam*, une insertion de l'élément *P* dans la région promotrice du cluster entraîne une perte des longs ARN précurseurs et l'élimination des *piARN* dérivés de ce cluster (Brennecke et al., 2007; Malone and Hannon, 2009). Le locus *flam* est l'exemple le mieux caractérisé de clusters de *piARN* primaires. Il s'exprime exclusivement dans les cellules folliculaires et contient plusieurs copies d'ET (*gypsy*, *Idefix*, *ZAM*) dont l'inactivation entraîne la surexpression de ces familles (Desset et al., 2008; Desset et al., 2003; Mevel-Ninio et al., 2007; Prud'homme et al., 1995). Les clusters de *piARN* constituent le recueil de l'information qui permet de déterminer la maturation d'un long transcrit précurseur en *piARN*. Ceux exprimés dans la lignée germinale sont particulièrement enrichis en ET. Ces derniers y sont présents sous forme complète, mais on les trouve généralement sous forme tronquée. Au niveau de certains clusters, plus de 90 % des *piARN* produits s'alignent avec des séquences d'ET (Aravin and Hannon, 2008; Aravin et al., 2008; Brennecke et al., 2007). Les clusters étant situés dans des régions riches en hétérochromatine, où les taux de recombinaisons sont faibles, on peut penser que lorsqu'un ET envahit une nouvelle espèce, il lui suffit de s'insérer dans un cluster préexistant pour acquérir la capacité à s'autoréguler. Il a été montré dans différentes espèces de drosophiles, que la localisation génomique des clusters était conservée mais qu'en revanche, leur contenu évoluait très rapidement (Malone and Hannon, 2009). Une question inexplorée est de savoir si l'évolution des clusters de *piARN* se fait de manière stochastique ou bien s'il existe un mécanisme actif permettant de nouvelles insertions spécifiquement dans ce type de régions.

Les raisons mécanistiques par lesquelles les *piARN* primaires sont produits, sont à ce jour peu décrites. On sait en revanche que l'inactivation de Piwi chez *D. melanogaster* (Mili et Miwi chez la souris) entraîne l'élimination du « pool » d'ARN primaires, ce qui ne signifie pas qu'elle soit directement impliquée dans l'initiation du processus. Concernant la maturation des *piARN*, on sait aujourd'hui que leur taille dépend majoritairement du membre de la famille PIWI avec lequel ils sont associés. On sait par exemple que les ARN d'une taille de 25 nucléotides sont préférentiellement associés à Piwi, ceux de 24 nucléotides à Aub et ceux de 23 nucléotides à Ago3. Cela suggère fortement que ces protéines agissent directement pour déterminer la taille du *piARN*.

De nombreuses autres protéines impliquées dans la biogenèse primaire ont été décrites chez *D. melanogaster*. On retrouve des protéines à domaine hélicase telles que Armitage et la famille des protéines Yb, la nucléase Zucchini, ainsi que la grande famille des protéines TUDOR (voir partie 3A). Des mutations dans ces protéines altèrent complètement la production des *piARN* primaires dans les cellules folliculaires (somatiques) de l'ovaire (Haase et al., 2010; Nishida et al., 2007; Olivieri et al., 2010; Qi et al., 2011). Par

ailleurs, les protéines Zucchini et Armitage semblent aussi responsables de l'accumulation de *piARN* primaires dans les cellules germinales de l'ovaire.

Les cellules folliculaires de l'ovaire chez *D. melanogaster*, sont un modèle pour l'étude de la biogenèse primaire puisque les *piARN* produits dans ces tissus, le sont uniquement *via* ce mécanisme. Il en est de même dans les testicules de souris, au niveau des spermatocytes qui entrent en méiose (Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006; Malone and Hannon, 2009; Saito et al., 2009). Cependant, d'autres types cellulaires, comme les cellules germinales de l'ovaire chez *D. melanogaster* et les spermatogonies pré-méiotiques chez la souris, présentent un autre mécanisme, le cycle ping-pong, qui permet l'amplification de séquences spécifiques produites par le mécanisme de biogenèse primaire.

B - La boucle d'amplification ping-pong

Chez *D. melanogaster*, les *piARN* primaires associés à Aub et Piwi sont pour la plupart en antisens par rapport à l'ARNm de l'ET (Brennecke et al., 2007; Nishida et al., 2007; Saito et al., 2006; Yin and Lin, 2007). Les *piARN* associés à Ago3, sont quant à eux, sont orientés dans le même sens que l'ARNm. Contrairement aux *piARN* pris en charge par Aub et Piwi, les *piARN* associés à Ago3 ne présentent pas le fort biais d'Uridine en position 1. Ils possèdent en revanche un fort biais en adénosine à la position 10. Une fraction significative des *piARN* associés à Ago3 et ceux associés à Aub ont la capacité de s'apparier sur 10 paires de base, entre les positions 1 et 10 (Brennecke et al., 2007; Gunawardane et al., 2007). Il a été montré par ailleurs, que les protéines PIWI avaient une activité de coupure au niveau du brin cible, précisément entre le 10^e et 11^e nucléotide relativement aux petits ARN guides (Gunawardane et al., 2007).

L'ensemble de ces observations a permis de proposer un modèle de synthèse : le modèle « ping-pong » dans lequel les protéines de la famille PIWI amplifient les *piARN* de manière cyclique grâce à leur activité de coupure. Ce modèle suppose dans un premier temps, qu'un *piARN* primaire (1U) pris en charge par Aub, cible un ARNm complémentaire pour le cliver, et ainsi générer un *piARN* secondaire possédant un A en position 10. L'activité enzymatique qui génère l'extrémité 3' des *piARN* secondaires est toujours inconnue. Cependant une fois qu'Ago3 est chargée par un *piARN* secondaire, ce duplex est à son tour capable de reconnaître et de cliver un transcrit complémentaire. A termes, cela permet de générer de nouveaux *piARN* qui correspondent exactement au *piARN* primaire initial et en conséquence, permet de réinitialiser un nouveau cycle d'amplification lorsque ceux-ci seront pris en charge par Aub (Brennecke et al., 2007; Gunawardane et al., 2007). Dans les testicules de *D. melanogaster*, la grande majorité des *piARN* retrouvés ne dérivent pas d'ET mais d'un allèle retrouvé sur le chromosome Y (*Suppressor of Stellate*) qui inhibe l'expression du gène *Stellate* (localisé sur le chromosome X). Ces *piARN* ne présentent pas de signature ping-pong et ne sont vraisemblablement pas amplifiés par un processus de biogenèse secondaire (Aravin et al., 2006). Le modèle ping-pong suppose donc que les protéines homologues Aub et Ago3 possèdent des fonctions distinctes au sein du cycle d'amplification ping-pong, ce qui explique leur rôle non-redondant dans la répression d'un même groupe d'ET.

Chez la souris, on observe le même phénomène pour les homologues Mili et Miwi2, c'est-à-dire qu'un même groupe d'ET ne réagit pas de la même façon en contexte mutant (Aravin and Hannon, 2008). Ici, le système d'amplification semble inversé en comparaison au modèle ping-pong chez *D. melanogaster*. En effet, la majorité des *piARN* primaires sont orientés dans le même que l'ARNm des ET tandis que les *piARN* secondaires sont majoritairement orientés en antisens. Dans les deux modèles, *D. melanogaster* et chez la souris, les données génétiques supportent fortement que l'établissement d'un processus d'amplification de type ping-pong nécessite l'intervention de deux homologues membres de la famille PIWI. Les mutants Mili abolissent la production de *piARN* primaires avec pour conséquence l'absence de *piARN* chargés par Mili2 ; chez les mutants Ago3, le nombre de *piARN* associés à Aub diminue très largement de même que l'orientation en antisens est moins marquée (Li et al., 2009). On ne peut cependant exclure l'hypothèse qu'un seul membre de la famille PIWI puisse à lui seul assurer cette double fonction (on parle de ping-pong homotypique). Cela a notamment été montré pour les *piARN* exprimés dans les spermatogonies chez la souris, où Mili est capable d'initier un cycle ping-pong sans son partenaire Miwi2 (Aravin and Hannon, 2008), de la même façon, chez les spermatocytes des drosophiles où Aub fonctionne sans son homologue Ago3 (Kai et al., 2014).

Il semble que le système ping-pong intervient pour affiner les populations de *piARN* de façon à ce que les ET les plus actifs soient ciblés préférentiellement (Aravin et al., 2007; Brennecke et al., 2007). En effet, on sait que les *piARN* sont amplifiés uniquement lorsque les transcrits sens et antisens sont présents dans la même cellule. La plupart des *piARN* dérivés des clusters sont produits en antisens et nécessitent par conséquent la présence d'un ET actif produisant des ARNm sens pour initier un cycle. Sous l'hypothèse qu'un ET présent en plusieurs centaines d'exemplaires est en moyenne plus transcrit qu'une copie unique, ce dernier va naturellement produire une plus grande quantité de matériel nécessaire à son inhibition et donc se trouver surreprésenté dans une population de *piARN*.

3- Les effecteurs protéiques de la voie des *piARN*

A - Les protéines à domaine *TUDOR*

Au niveau mécanistique, les protéines de la famille PIWI, de part leur association avec les petits ARN et leur activité endonucléase, suffisent à expliquer l'émergence des *piARN* dans lignée germinale. Cependant, les études génétiques menées chez la souris et la drosophile montrent que l'efficacité du processus de répression par la voie des *piARN* dépend d'un grand nombre d'effecteurs protéiques. Parmi ces effecteurs plusieurs ont la capacité d'interagir directement avec les protéines PIWI. L'analyse biochimique des modifications post-traductionnelles des protéines PIWI a montré que plusieurs résidus de type arginine étaient très conservés et diméthylés. Ces résidus sont en effet essentiels pour générer une plateforme d'interaction avec des protéines contenant un domaine *TUDOR* (Siomi et al., 2010; Vagin et al., 2009). Les analyses biochimiques des protéines

co-purifiées avec les protéines PIWI chez *D. melanogaster* et la souris ont montré que les protéines PIWI s'associaient avec des protéines de type arginine N-méthyltransférase (PRMT5 ou Capsuleen chez *D. melanogaster*), qui permettent la méthylation des résidus Arginine (Kirino et al., 2009; Nishida et al., 2009; Vagin et al., 2009). Cette méthylation a lieu de concert avec une protéine du méthylosome (*MAP50* ou *Valois* chez *D. melanogaster*) et elle est observée chez plusieurs espèces non apparentées (Chen et al., 2009; Nishida et al., 2009; Reuter et al., 2009; Vagin et al., 2009). Ces résidus arginines sont en revanche absents dans les protéines du clade AGO suggérant que cette étape de méthylation est caractéristique de la voie des *piARN*. Des études de protéomiques sur les complexes protéiques PIWI ont révélé que plusieurs protéines, contenant un ou de multiples domaines *TUDOR*, étaient présentes et connues pour s'associer avec la structure diméthylée des protéines PIWI (Chen et al., 2009; Reuter et al., 2009; Vagin et al., 2009). Chez *D. melanogaster*, la protéine Tudor (Tud) est un facteur indispensable à la gamétogenèse qui contient 11 domaines *TUDOR*. Il a été montré que la protéine s'associait directement avec les protéines de la famille PIWI, Aub et Ago3 (Kirino et al., 2009; Nishida et al., 2009), suggérant encore une fois leur rôle essentiel dans la biogenèse secondaire des *piARN*. Plusieurs autres protéines indispensables à la régulation par les *piARN* possèdent un domaine *TUDOR* unique, c'est le cas de *Spindle-E*, des protéines *Yb*, de *Krimper*, *Qin*, *PAPI* et *Tejas* (Ku and Lin, 2014). Les protéines contenant plusieurs domaines *TUDOR* peuvent potentiellement interagir simultanément avec d'autres protéines et donc fonctionner comme une plateforme d'assemblage/repliement pour la formation d'un complexe multi-protéique.

Cette idée est particulièrement intéressante dans le cadre du mécanisme ping-pong qui requiert la grande proximité entre deux protéines de la famille des PIWI. On sait grâce à une étude cristallographique, que la protéine Tud s'associe simultanément à plusieurs protéines Aub (Liu et al., 2010). On sait aussi que des mutants de protéines à domaine *TUDOR* perturbent le cycle d'amplification ping-pong sans l'abolir complètement cependant (Nishida et al., 2009; Reuter et al., 2009; Vagin et al., 2009). Plusieurs arguments indépendants montrent que le répertoire de *piARN* est drastiquement différent après la mutation d'une protéine contenant un domaine *TUDOR*. En définitive, il semble que l'ensemble de ces protéines à domaine *TUDOR*, possèdent des rôles redondants concernant leur interaction avec le duplex de protéines PIWI, ce qui expliquerait en partie pourquoi les phénotypes associés ne soient pas totalement délétères.

B - Une large diversité d'interactants protéiques

Plusieurs autres protéines ne contenant pas de domaine *TUDOR* sont essentielles à la régulation par les *piARN*. Parmi celles-ci la protéine Armi (une hélicase ARN) et son homologue (Mov10L1) chez la souris (présent chez tous les mammifères) sont essentielles à la biogenèse primaire chez *D. melanogaster* (Lim and Kai, 2007; Olivieri et al., 2010; Zheng et al., 2010). Au niveau mécanistique, on ignore toujours si le domaine hélicase associé à Armi possède une quelconque activité et quel est son rôle précis dans la biogenèse primaire. Armi, semble néanmoins affecter la biogenèse secondaire, puisque les mutants, montrent une diminution des *piARN* pris en charge par Piwi et Aub.

Une autre protéine, Vasa, chez *D. melanogaster* (et son homologue Mvh chez les mammifères), possède un domaine hélicase. Il s'agit d'un composant essentiel du complexe nuage qui se situe à la périphérie du noyau des cellules germinales de l'ovaire (décrit dans la partie suivante, 4A). Au niveau structural, Vasa possède une plateforme de liaison (résidus arginine méthylés) la famille des protéines à domaine *TUDOR*, tout comme les protéines de la famille PIWI. De concert avec Aub et Ago3, Vasa s'accumule au niveau du complexe nuage et, en contexte mutant, y déstabilise le recrutement de ces deux protéines. Au niveau phénotypique, cela entraîne une réduction significative des *piARN* impliqués dans le cycle ping-pong (Kuramochi-Miyagawa et al., 2010; Liang et al., 1994; Lim and Kai, 2007; Nagao et al., 2010; Toyooka et al., 2000; Vasileva et al., 2009). La biogenèse secondaire a lieu dans le cytoplasme et, une étape importante du processus, est l'export de l'ARN précurseur à l'extérieur du noyau. On pense que Vasa est essentielle à l'assemblage du complexe nuage et par conséquent, à la mise en place des structures péri-nucléaires associées à la biogenèse, comme les structures d'exportation par exemple.

La protéine Maelstrom, également présente dans le complexe nuage, semble être impliquée dans la voie des *piARN*, spécifiquement dans le noyau. Il semble que, de concert avec Piwi, cette protéine va permettre la répression transcriptionnelle des ET (Sienski et al., 2012). Cependant, on ignore les interactions moléculaires de *Maelstrom* avec Piwi. De la même façon, des protéines comme Squash (une nucléase) n'affecte pas les populations de *piARN* mais ont un effet fort sur la dérégulation transcriptionnelle des ET. De manière générale, c'est le cas de nombreuses de protéines interagissant avec les protéines PIWI. Celles-ci affectent le processus de répression des ET mais on ignore à ce jour, les bases de leur fonctionnement moléculaire.

4- Régionalisation cellulaire au cours de la biogenèse des *piARN*

A - Le complexe nuage

Les protéines de la famille PIWI sont localisées à la fois dans le cytoplasme, le noyau et à l'interface des deux compartiments au niveau du complexe appelé nuage. Dans le cytoplasme, les protéines PIWI et leurs interactants colocalisent dans différents granules cytoplasmiques présents dans les cellules germinales des métazoaires. Malgré la présence de Piwi ou Miwi2 dans le noyau, la majorité des observations montrent que la production et le chargement des *piARN* a lieu dans ce type de granules (Aravin et al., 2009; Olivieri et al., 2010; Saito et al., 2009; Saito and Siomi, 2010). De plus, les granules cytoplasmiques qui contiennent les composants de la voie des *piARN* sont majoritairement situés à proximité du noyau, suggérant leur implication dans la reconnaissance du précurseur de *piARN* et son export vers le cytoplasme (Kojima et al., 2009; Nagao et al., 2010; Nishida et al., 2009; Nishida et al., 2007; Saito and Siomi, 2010; Shoji et al., 2009). Les granules germinaux correspondent au complexe nuage présent dans l'ovaire chez la drosophile (dans les cellules nourricières). Les composants du complexe nuage et notamment les protéines PIWI chargées par les *piARN*,

sont transmis des cellules nourricières vers l'ovocyte au niveau postérieur (Anne and Mechler, 2005; Houwing et al., 2007; Kirino et al., 2009). Certains de ces composants donnent naissance au plasma germinatif qui permet d'induire la formation des cellules germinales primordiales chez l'embryon (Illmensee and Mahowald, 1974). Le nuage contient donc la plupart des protéines requises pour la répression *via* les *piARN* (Spn-e, Mael, Vasa, Krimp, Tejas, Tud), dont une partie est transmise maternellement à la descendance. Comme décrit précédemment, lorsque ces protéines sont déficientes, on assiste à l'élimination des *piARN*, à une modification du répertoire standard ou bien l'absence de localisation des protéines de la famille PIWI. Pour ces raisons, il est supposé que le nuage est une structure mise en place autour du noyau des cellules germinales, essentielle à la biogenèse des *piARN*, la reconnaissance et la destruction de leurs cibles (Lim and Kai, 2007).

B - Les granules Yb des cellules somatiques de l'ovaire

Chez *D. melanogaster*, on peut étudier le processus de biogenèse primaire et la répression des ET, soit dans les cellules somatiques de l'ovaire, soit par le biais de lignées cellulaires OSC, dérivées des cellules folliculaires (Malone and Hannon, 2009; Olivieri et al., 2010; Saito et al., 2009; Saito and Siomi, 2010). A la différence des cellules germinales de l'ovaire, ces cellules ne présentent pas de complexe nuage et aucun des marqueurs du nuage n'y est exprimé (Vasa, Aub et Ago3). De surcroît, les protéines généralement requises pour produire des *piARN* dans les cellules germinales, ne sont pas indispensables au processus de biogenèse primaire dans les cellules folliculaires (Spn-e, Vasa, Mael, Tud et Krimp) suggérant que les machineries de production diffèrent entre ces types cellulaires. Cependant, il a été observé dans le cytoplasme des cellules folliculaires, un type de structure granulaire, potentiellement proches de la structure granuleuse du nuage, appelées « corps Yb » (« *Yb-bodies* ») (Olivieri et al., 2010; Qi et al., 2011; Saito and Siomi, 2010). Ces granules ne colocalisent pas avec d'autres types de granules (P-bodies ou U-bodies) connus pour être impliqués dans le métabolisme des complexes des protéines ribonucléaires (snRNP). L'analyse génétique des corps Yb a montré que les protéines Piwi, Arm, Zuc, et FS(1)Yb y étaient recrutées et requises pour la production et l'accumulation des *piARN* dans la cellule (Olivieri et al., 2010; Qi et al., 2011; Saito and Siomi, 2010). La protéine *FS(1) YB* est essentielle à la formation des corps Yb et nécessaire à la localisation d'Arm et Piwi dans ces granules. En l'absence de l'endonucléase Zuc, les protéines Arm et Piwi sont localisées uniquement dans les corps Yb et ne peuvent plus pénétrer à l'intérieur du noyau, ce qui coïncide avec l'élimination des *piARN* primaires. Enfin, en présence d'une déficience de la protéine Arm, les protéines Piwi ne sont plus chargées par les *piARN* et ne localisent plus dans le noyau, suggérant que le chargement de Piwi est requis pour son entrée dans le noyau. L'ensemble de ces résultats, suggère donc que les corps cytoplasmiques Yb soient le type de structures au sein desquelles les fonctions du complexe *piRISC* seraient contrôlées avant de permettre leur entrée dans le noyau. De manière générale, les corps granuleux sont des structures indispensables au maintien des mécanismes associés à la production des *piARN* que ce soit dans les cellules somatiques folliculaires ou dans les cellules germinales nourricières, bien que leur composition diffère entre les types cellulaires.

5- La co-évolution des mécanismes de régulation et le génome hôte

A - Un modèle analogue aux interactions hôtes-pathogènes ?

En tant que parasites génomiques, les ET et leurs interactions avec le génome hôte sont parfois décrites comme un processus de co-évolution entre l'hôte et ses pathogènes. Les effets délétères des ET sur la fitness de l'hôte suggèrent que ces interactions sont analogues à la course aux armements qui existe entre le système immunitaire du génome et ses pathogènes (Aravin et al., 2007; Kidwell and Lisch, 2001; Siomi et al., 2008). Lorsque les pathogènes sont transférés horizontalement, on assiste souvent à l'émergence de mutations avantageuses au niveau des allèles impliqués activement dans la suppression de l'infection par le pathogène (Elde et al., 2012). Si cette relation est vraie, on s'attend alors à observer au niveau moléculaire, les mêmes signatures évolutives pour les gènes associés à la répression des ET. Cependant, l'extension de ce modèle est un peu plus compliquée du point de vue des ET car la plupart des familles d'ET sont héritées verticalement contrairement aux organismes pathogènes qui apparaissent spontanément au cours de la vie de l'individu. En effet, lorsque les individus d'un groupe s'accouplent de manière aléatoire, les taux de recombinaisons ont pour effet de limiter l'association entre une mutation sélectionnée pour supprimer l'activité d'un ET et la perte d'activité de l'élément (Charlesworth and Langley, 1986). Ce mode de transmission favorise peu les avantages sélectifs au niveau des gènes répresseurs pour une famille d'éléments en particulier.

Cependant, dans de rares cas, les ET sont transmis d'une espèce à l'autre par transferts horizontaux (Oliveira de Carvalho et al., 2004; Schaack et al., 2010; Silva et al., 2004). Dans ces cas précis, l'impact évolutif des ET sur le génome hôte est analogue à celui d'un pathogène non génomique comme les virus par exemple. Au cours de l'expansion d'une nouvelle famille d'ET, les taux de transpositions sont particulièrement élevés (Good et al., 1989; Kidwell et al., 1988; Le Rouzic and Capy, 2005). Ce phénomène est souvent associé à une forte incidence d'insertions délétères et de stérilité impliquant par conséquent de fortes pressions de sélections sur le génome hôte, suggérant l'émergence éventuelle d'une réponse adaptative. L'exemple le mieux étudié est celui de l'élément *P* chez *D. melanogaster* transféré horizontalement depuis *D. willistoni* (35 millions d'années de divergence), il y a environ 60 ans (Anxolabehere et al., 1988; Brookfield et al., 1984; Clark et al., 1994; Daniels et al., 1990). L'évolution des gènes associés à la régulation par les *piARN* a été comparée entre les souches qui présentent l'élément et celles qui en sont exemptes (Langley et al., 2012). L'impact de la sélection sur ces gènes est faible et ces derniers ne semblent pas avoir été affectés par le transfert horizontal de l'élément *P*. Cependant, indépendamment du transfert horizontal, il semble que l'ensemble des gènes associés à la répression des ET *via* les *piARN* soit soumis de façon récurrente à la fixation de mutations adaptatives, comme on l'observe pour les gènes impliqués dans la résistance aux pathogènes (Langley et al., 2012). Cette observation suggère que ce n'est pas une famille unique qui est responsable de l'évolution rapide de ces gènes. Il semble au contraire, que ces gènes sont contraints par la charge globale en ET dans le génome.

Au niveau du mécanisme, une analogie avec le système immunitaire a également été proposée. Les deux étapes de la biogenèse des *piARN* (le processus de genèse primaire et la boucle d'amplification secondaire) sont directement impliquées dans la mise en place de défenses contre l'invasion des ET. La biogenèse primaire peut être comparée à la production initiale du répertoire « hypervariable » d'anticorps et la boucle d'amplification à la sélection clonale des lymphocytes producteurs d'anticorps pendant la réponse immunitaire acquise. L'importance de la boucle d'amplification dans l'acquisition d'un mécanisme de défense contre les ET, a été mise en évidence par l'étude des phénomènes de dysgenèse des hybrides chez *D. melanogaster*. Ce phénomène est observé lorsque l'on croise deux souches différentes qui donnent une descendance stérile et présentant des gonades complètement atrophiées. Ce phénotype a souvent été associé à l'activation de certaines familles de transposons comme les éléments *P*, *I* et *hobo*. La particularité de ce syndrome est qu'il est observé uniquement quand les mâles, contenant des copies actives, sont croisés avec les femelles dépourvues de copies actives, tandis que le croisement réciproque ne présente aucun phénotype anormal. Au regard de leur génotype, les individus hybrides sont identiques et par conséquent un facteur cytoplasmique maternellement transmis est vraisemblablement à l'origine de ce syndrome (Figure 27). Il a été montré que l'activation des ET avait pour cause l'absence de *piARN* spécifiques, normalement déposés maternellement dans l'ovocyte pour réprimer les ET transmis par le père. Le dépôt maternel des complexes *piRISC* chargés en *piARN* fournit à l'embryon le matériel initial nécessaire à initier la boucle d'amplification ping-pong et donc acquérir une « immunité » contre les effets délétères des ET. Par ailleurs, le syndrome de dysgenèse hybride s'estompe au cours de la vie de l'individu, on observe notamment une restauration de la fertilité chez les femelles au cours du temps. Cette restauration est observée tant au niveau du nombre d'œufs pondus par la femelle, qu'au niveau du pourcentage d'éclosion de ces œufs (Khurana et al., 2011). Il a été montré que l'absence de *piARN* transmis par la mère conduisait dans un premier temps, à la forte mobilisation de l'élément *P*, à de nombreuses cassures double brin et à l'arrêt de la prolifération des cellules souches (par un checkpoint cellulaire). Au cours du temps, l'élément *P* continue sa mobilisation et finit par produire *de novo* des *piARN* de manière à ré-initier la biogenèse secondaire et restaurer en partie, la fertilité, notamment dans les cellules ayant acquises une insertion spécifique dans un cluster de *piARN*. Cet exemple montre clairement qu'au cours de la vie d'un individu se mettent en place de nouveaux allèles de résistance contre les ET *via* de nouvelles insertions dans les clusters de *piARN*. Cela suggère de surcroît que les clusters varient rapidement en terme de contenu et qu'il s'agirait de réservoirs de défense anti-ET régulièrement actualisés par de nouvelles insertions d'ET, notamment les ET mobilisés à forte fréquence dans la lignée germinale.

B - La divergence des protéines associées à la régulation par les *piARN* : une diversification extrêmement rapide

L'adaptation de l'hôte à de nouveaux ET qui envahissent le génome est supposée acquise lorsque celui-ci s'insère au sein d'un cluster de *piARN* de manière à produire et transmettre le matériel « immunitaire » nécessaire à sa répression dans la descendance et sur plusieurs générations. Comme décrit précédemment, ce

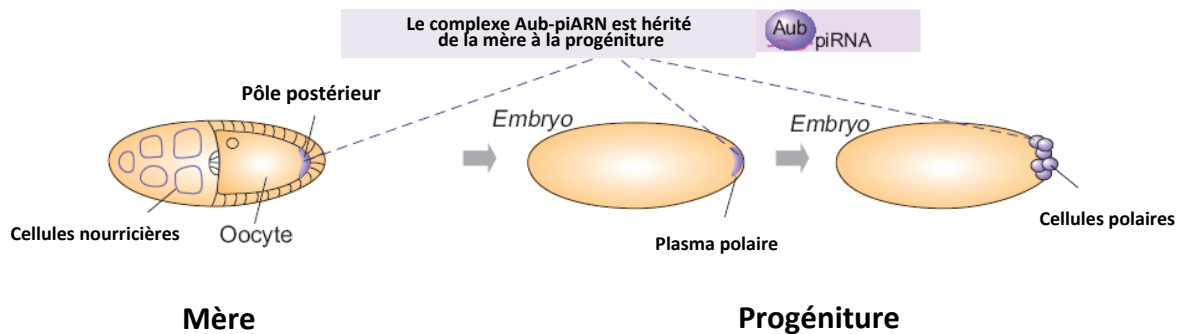


Figure 27. Héritage maternel des *piARN* associés à Aub

Aub est accumulé au niveau du pôle postérieur de l'oocyte et peut être détecté dans les cellules embryonnaires du pôle suggérant que les *piARN* sont hérités dans les cellules germinales progénitrices de la mère sous la forme d'un complexe Aub/*piARN*. Seul les *piARN* antisens sont hérités en raison de l'absence d'Ago3 au niveau du pôle postérieur des oocytes. Adapté de (Senti and Brennecke, 2010).

constat semble clair entre deux souches (populations) de la même espèce. Ce modèle d'adaptation aux nouvelles insertions est particulièrement séduisant puisqu'il est indépendant des variations associées aux gènes de la régulation des *piARN*. Si ce système est effectivement auto-suffisant et que la répression des ET dépend exclusivement du *pool* de *piARN* hérités maternellement, on s'attend à observer les mêmes caractéristiques phénotypiques dans le cas d'un croisement dysgénique intraspécifique et interspécifique. Dans les deux cas, on s'attend à ce que les ET hérités par la mère ne soit pas dérprimés chez l'hybride et qu'à l'inverse, les familles d'ET uniquement présentes chez le père soient fortement dérprimées. Néanmoins, lorsque l'on croise les espèces sœurs *D. simulans* et *D. melanogaster*, la situation dans l'hybride interspécifique est caractérisée par une dérpression globale de l'ensemble des familles d'ET, indépendamment de l'origine du *pool* initial de *piARN*. En effet, la dérpression des ET n'est pas associée à l'absence de *piARN* chez la mère et touche de la même façon l'ensemble des familles d'ET. En fait, la situation chez l'hybride interspécifique est très proche des phénotypes observés pour les mutants des effecteurs protéiques de la voie des *piARN* : une dérpression globale des ET ainsi qu'une production aberrante de *piARN*. De plus, au niveau cytologique dans les cellules germinales, on assiste à une perte de la localisation des protéines Aub et Ago3 au niveau du complexe nuage (Figure 28). L'hypothèse avancée pour expliquer un tel phénotype est que depuis les 6 millions d'années de divergence entre *D.simulans* et *D.melanogaster*, les protéines associées à la régulation des ET se sont adaptées au contenu en ET dans leur génome respectif et que cette adaptation est si rapide que les protéines d'une espèce se comportent comme des allèles non-fonctionnels dans l'autre espèce (Kelleher et al., 2012).

Ces observations suggèrent que dans la lignée germinale, il existe des conflits entre le génome hôte et les ET qui conduisent à une « course aux armements » (Charlesworth and Langley, 1989), au même titre que pour les virus (Marques and Carthew, 2007; Obbard et al., 2006). Cette hypothèse est soutenue par plusieurs

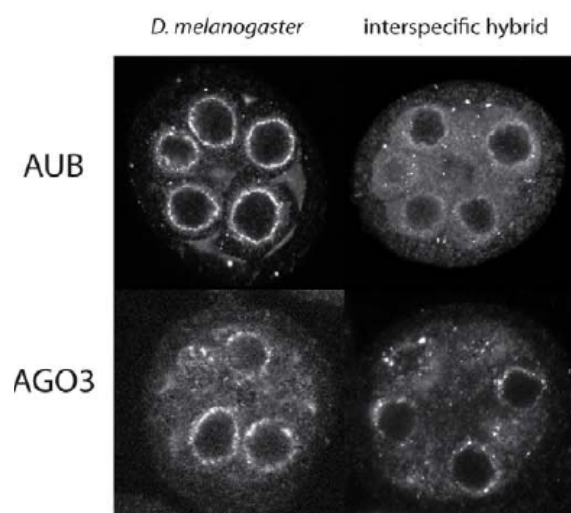


Figure 28. Localisation d'Aubergine et d'Ago3

La figure montre la localisation des protéines Aub et Ago3 au niveau du complexe nuage dans les cellules nourricières d'ovaires, chez *D. melanogaster* (à gauche) et chez l'hybride interspécifique (à droite). Chez *D. melanogaster*, les deux protéines se localisent correctement autour du noyau au sein du nuage de manière à initier la production de *piARN* secondaires. Chez l'hybride interspécifique, les deux protéines ne sont plus localisées au niveau du complexe nuage. Ce phénotype est associé à la dérégulation globale des ET chez l'hybride interspécifique. L'hypothèse soutenue par les auteurs (Kelleher *et al.* 2012) est que ces protéines ont rapidement divergées entre *D. melanogaster* et *D. simulans* (moins de 5 millions d'années), à tel point que ces dernières sont incapables d'effectuer leur rôle chez l'hybride. Issu de (Kelleher and Barbash, 2013).

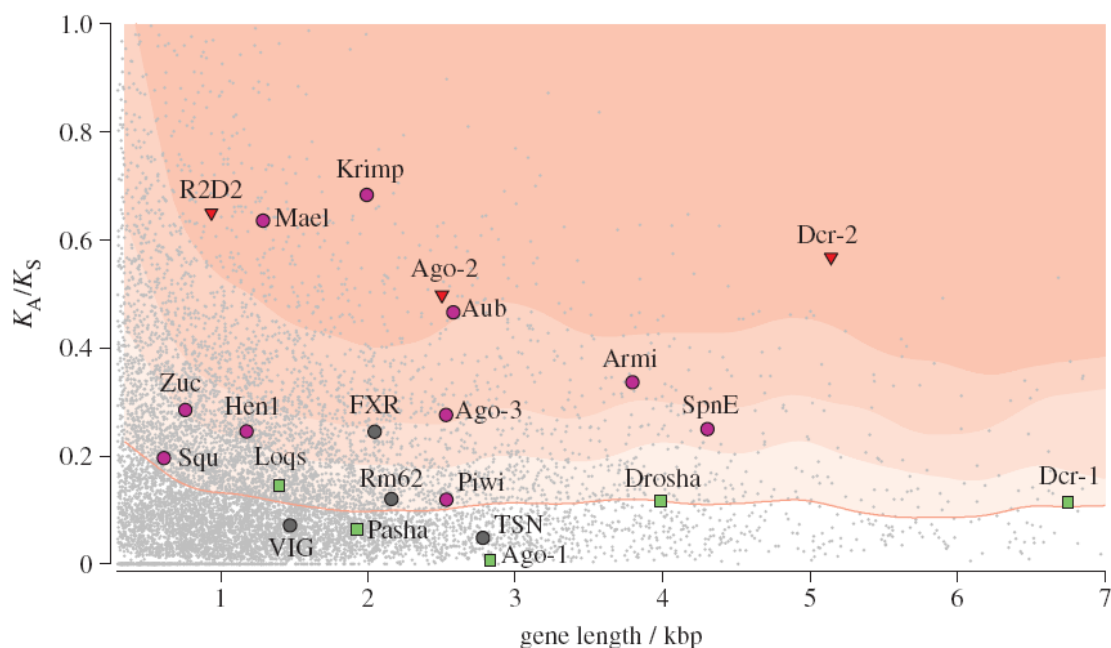


Figure 29. Taux d'évolution des gènes associés au mécanisme d'ARN interférence chez la drosophile

Le taux d'évolution des protéines (K_A , sites non synonymes) relatif au taux d'évolution neutre (K_S , sites synonymes) entre *D. melanogaster* et *D. simulans* est représenté en fonction de la longueur des gènes. Le K_A/K_S est calculé selon la méthode de Li et al (1985). Tous les 10 581 gènes orthologues dont les séquences sont disponibles sont représentés en gris. Les 22 gènes de la voie des ARN interférents sont représentés par des ronds, des carrés et des triangles. La médiane (ligne rouge) et les 75^{ème}, 85^{ème}, 95^{ème} percentiles (fond colorés) sont représentés pour 380 gènes sauvages. Les gènes associés à l'ARN interférence évoluent significativement plus rapidement que la moyenne des gènes sauvages (Wilcoxon signed-rank test, $p < 0.01$). Six d'entre eux sont situés dans les 5% de gènes évoluant le plus rapidement (*R2D2*, *maelstrom*, *krimper*, *Ago-2*, *aubergine* and *dcr-2*). 12 sont situés au-dessus de 50% des gènes. La voie des *piARN* (ronds violets) et des *siARN* (triangles rouges) évoluent plus rapidement que les autres gènes ($p < 0.001$ et $p < 0.01$, respectivement), contrairement aux gènes de la voie des *miARN* (carrés verts) ($p < 0.34$). Issu de (Obbard et al., 2009a).

études portant sur le polymorphisme de séquence des gènes associés à la régulation des *piARN*. L'étude des 12 génomes de drosophile a montré que les gènes *krimper* et *spindle-E*, sous l'action de la sélection naturelle, avaient des taux d'évolution moléculaire particulièrement élevés (Heger and Ponting, 2007). Chez *D.simulans*, les gènes *maelstrom*, *krimper* et *aubergine* font parti des 5 % de gènes évoluant le plus rapidement. De même, un grand nombre de mutations adaptatives ont été observées pour les gènes *armitage*, *aubergine*, *piwi*, *maelstrom* et *spindle-E* (Figure 29) (Obbard et al., 2009b). Chez *D. melanogaster*, on observe des marques de sélection positives similaires pour les gènes associés à la régulation par les *piARN* (e.g. *aub*, *krimp*, *squ,zuc*) (Berry et al., 2009; Kolaczowski et al., 2011; Langley et al., 2012; Pane et al., 2007).

Ces observations montrent que les protéines associées à la régulation des *piARN* sont régulièrement sous pressions de sélection diversifiantes. Cependant, le lien avec le contenu en ET dans le génome n'est pas direct. Une étude a montré que l'abondance des ET dans les génomes du genre *Drosophila* n'était pas corrélée à une augmentation du taux de substitutions non-synonymes pour les gènes de la machinerie des *piARN* (Blumenstiel, 2011). Il semble au contraire, que l'augmentation des ET dans les génomes soit corrélée positivement à un fort biais de codons qui vise à optimiser l'efficacité traductionnelle de la machinerie de régulation. Ce modèle suggère que la course aux armements entre ET et la machinerie de régulation de l'hôte serait dans un premier temps limitée par la nécessité de produire plus de protéines réprimant l'activité transcriptionnelle des ET.

Ces données au niveau de la séquence représentent une approximation quant à l'activité transcriptionnelle réelle de ces gènes. Il a été montré par ailleurs qu'au sein d'une même espèce, l'activité transcriptionnelle des gènes associés à la régulation par les *piARN* était extrêmement variable (Fablet et al., 2014). En effet, une étude menée dans 10 souches de *D. simulans* d'origines géographiques différentes montre que les niveaux de variations transcriptionnelles des 7 gènes de la machinerie des *piARN* testés sont très largement supérieurs à un pool de gènes indépendants. En association avec le taux d'évolution des séquences de ces gènes, il semble que les taux de mutations non-synonymes soient aussi significativement supérieurs à ceux des gènes non associés à la machinerie des *piARN*.

Le lien direct entre l'évolution de la machinerie des *piARN* et le contenu en ET dans le génome est à ce jour controversé et nécessite de plus larges explorations. Ce lien est multifactoriel et a pu être modulé à différents niveaux au cours de l'évolution. Il y a des contraintes au niveau populationnel, qui déterminent la balance entre dérive et sélection naturelle, tant pour les ET que pour les gènes associés à leur régulation. Il y a des contraintes génomiques comme par exemple le *pool* d'ET présents dans le cluster de *piARN*, des contraintes développementales qui dépendent par exemple du *pool* de *piARN* transmis par la mère, et de manière plus globale, des contraintes environnementales qui peuvent affecter l'histoire des ET, la machinerie des *piARN* et l'ensemble du génome à de multiples niveaux.

Problématique

La trajectoire évolutive des ET dans le génome est souvent difficile à prédire, ne serait-ce qu'au niveau d'une seule famille. Néanmoins, certaines forces majeures, telles que la dérive ou la sélection naturelle, s'appliquent à l'ensemble des lignées tant au niveau des génomes qu'au niveau des ET. Ces forces façonnent les interactions entre le génome et les ET au cours d'une histoire commune qui s'étale sur différentes échelles de temps évolutifs. Comparer les compositions en ET chez différentes espèces, donne une vision sur le long terme caractérisée par une très grande hétérogénéité tant sur le contenu que sur la diversité. Nous avons vu par ailleurs, qu'une transition majeure, comme l'apparition du sexe, a un impact direct sur le maintien des ET dans le génome. Par exemple, l'émergence des chromosomes sexuels conduit de manière convergente, à une forte accumulation en ET dans certaines régions du génome et à leur rapide dégénérescence fonctionnelle. Ce type d'approche rend néanmoins difficile l'établissement d'un lien direct entre l'histoire d'une lignée d'ET et les mécanismes mis en place par le génome pour en limiter l'invasion. Chez les métazoaires sexués, le lieu clef des interactions entre le génome et les ET est la lignée germinale. Une forte activité des ET dans la lignée germinale est souvent associée à des phénotypes sévères, compromettant le succès reproducteur de l'individu et la survie de sa descendance. Cependant, différents mécanismes ont été sélectionnés pour limiter les effets délétères causés par les familles actives d'ET. Ces mécanismes de régulation font intervenir les ARN interférents, qui interagissent directement avec les ARNm des ET de manière à inhiber leur transcription. A ce jour, il reste de nombreuses zones d'ombres sur la manière par laquelle la régulation des ET dans la lignée germinale affecte leur dynamique dans le génome.

Mes travaux de thèse sont présentés en trois chapitres qui s'intéresseront au lien entre la régulation des ET par les voies d'ARN interférences et leur activité au niveau de l'individu, de la population et de l'espèce. Ces trois approches permettent de visualiser la relation entre ET et les mécanismes mis en œuvre par le génome à différentes échelles biologiques et à différents moments au cours de l'invasion d'un ET.

Au niveau de l'individu, nous étudierons le cas d'un ET récemment introduit dans le génome, de manière à comprendre les mécanismes mis en jeu au cours de l'invasion précoce. L'élément *mos1* (transposon de Classe II de la famille *mariner*) a été introduit par transgénèse, il y a plusieurs années chez *D. melanogaster*. L'élément s'est amplifié à hauteur de 20 copies dans le génome et demeure stable depuis. De manière à savoir si cet état d'équilibre est lié à l'établissement d'une régulation par les ARN interférents, nous avons étudié l'activité de l'élément dans des mutants des voies de régulation des *siARN* et des *piARN*. L'analyse génétique a été complétée par un séquençage du transcriptome afin d'identifier les populations de petits ARN dérivés de *mos1*. Par la suite, nous avons suivi l'évolution du nombre de copies sur plusieurs générations (en présence ou en absence de *piARN*) pour en comprendre l'impact sur la dynamique de l'élément. Enfin, nous avons caractérisé plusieurs insertions de manière à comprendre le contexte génomique dans lequel la régulation peut s'établir et se maintenir.

Au niveau populationnel, la question a été de savoir si les variations d'activité de l'élément *mos1* entre populations d'une même espèce, peuvent être associées au polymorphisme transcriptionnel des gènes de régulation par ARN interférence. Une analyse détaillée de l'activité de l'élément (activité transcriptionnelle, nombre de copies et capacité d'excision) a été menée sur un panel de 19 populations de l'espèce *D. simulans*. Ces populations se différencient par leur histoire démographique et ont colonisé de nouvelles niches environnementales à différents moments. Plusieurs paramètres environnementaux (*e.g.* la température, la latitude *etc.*), la distance d'invasion et les variations d'expressions des gènes de régulation par ARN interférences ont été testés et mis en relation avec les variations d'activité de *mos1*. L'ensemble des données est synthétisé sous la forme d'un scénario évolutif décrivant les facteurs impliqués dans les changements d'activité de *mos1* au cours du processus de colonisation.

Dans le troisième chapitre, nous décrivons le type de réponse globale apporté par les petits ARN, chez les deux espèces sœurs, *D. melanogaster* et *D. simulans* qui comportent un contenu et une activité en ET différentes. Une question importante posée dans ce chapitre est de savoir comment se comportent les petits ARN lorsque le génome est confronté à d'importantes vagues d'amplifications d'ET. Nous avons, dans un premier temps, étudié l'activité transcriptionnelle des ET spécifiques des drosophiles dans la lignée germinale des espèces sœurs *D. melanogaster* et *D. simulans*, via des données de transcriptomes. Nous avons ensuite caractérisé le degré d'amplification de chaque famille dans les génomes de références. Une fois l'activité des familles caractérisée dans chacune des espèces, nous avons, à l'aide de données de « deep-sequencing », établi les liens entre les populations de petits ARN et l'évolution de chacune des lignées d'ET actives dans la lignée germinale.

Au cours de ces trois chapitres, nous aborderons enfin une autre question fondamentale, relative à l'évolution des systèmes de régulation entre la lignée germinale mâle et la lignée germinale femelle. Nous reviendrons sur les différences majeures entre ces deux lignées, du point de vue de l'activité des ET, des populations de petits ARN et de l'activité des gènes associés à leur régulation. En effet, dans la littérature, l'ensemble des questions posées autour de la régulation des ET dans la lignée germinale se focalise sur l'ovaire. Peu d'études adressent cette question au niveau de la lignée mâle qui, pourtant, se comporte très différemment.

Résultats

Article n°1

La voie des *piARN* limite l'expansion de *mos1* dans la lignée germinale

Article n°1

La voie des *piARN* limite l'expansion de *mos1* dans la lignée germinale

Bastien SAINT LEANDRE, Isabelle CLAVEREAU, Emilie ROBILLARD, Pierre CAPY, Aurélie HUA-VAN

Introduction

Les éléments transposables (ET) sont des segments d'ADN répétés qui possèdent la capacité de se multiplier et de se propager dans les génomes hôtes. Néanmoins, lorsqu'un ET émerge dans un nouveau génome, ce dernier est confronté aux forces de sélection négatives et aux mécanismes de défense de l'hôte qui tendent à limiter sa transmission verticale et luttent contre leurs effets délétères. Les génomes eucaryotes ont acquis un certain nombre de mécanismes de défenses pour lutter contre l'invasion de séquences exogènes. Les voies de régulations associées à la production d'ARN interférents (ARNi), ont clairement démontré leur fonction de défense contre les effets néfastes des virus et des ET dans une grande variété d'organismes, incluant les nématodes (Sijen and Plasterk, 2003; Tabara et al., 1999), les drosophiles (Blumenstiel and Hartl, 2005; Brennecke et al., 2007), les plantes (Hamilton and Baulcombe, 1999; Ito et al., 2011; Zilberman et al., 2003) et les mammifères (Aravin and Tuschl, 2005; Aravin and Hannon, 2008).

Les voies de régulation à ARN sont toutes associées à la production de petits ARN (d'une taille comprise entre 20 et 30 nt), qui agissent comme des facteurs spécifiques capables d'inactiver l'expression de séquences homologues. Chez les animaux, la famille AGO et la sous-famille PIWI constituent les deux classes majoritaires de protéines conservées Argonautes. Ces protéines interagissent directement avec les ARNi pour lesquels on compte trois grandes classes. Les *siARN* (*short interfering RNA*) et les *miARN* (*microRNA*) ont une taille comprise entre 21 et 25 nt et interagissent directement avec les protéines de la famille AGO (e.g. Drosha et Dicer). Ces petits ARN sont générés *via* l'enzyme Dicer, à partir de précurseurs d'ARN double brins plus longs (e.g. <100 nt pour les *miARN*) qui dérivent, soit de séquences endogènes (présentes dans le génome l'hôte), soit de séquences exogènes (présentes uniquement dans le cytoplasme). Les petits ARN guides ainsi produits permettent l'appariement avec leurs ARNm homologues, de manière à cibler les complexes effecteurs RISC (dans le cytoplasme) et RITS (dans le noyau). Le cœur de chaque complexe contient les domaines PAZ et Piwi (hautement conservés) qui caractérisent les protéines Argonautes et permettent l'interaction avec les petits ARN. L'activité de clivage des complexes effecteurs est assurée par le domaine RNaseH des protéines AGO. Les protéines de la sous famille PIWI (Piwi, Ago3, Aub) interagissent quant à elles, spécifiquement avec la troisième classe de petits ARN, les *piARN* (*Piwi-interacting RNA*). Les *piARN* se différencient de par leur taille légèrement plus longue (23-31 nt). Bien qu'à ce jour le mécanisme exact par lequel ces *piARN* sont produits n'est pas complètement compris. Le modèle dominant suppose que les *piARN* sont produits à partir d'un petit nombre régions dans le génome appelées les clusters de *piARN*. La production des *piARN* est initiée au niveau de ces clusters sous la forme d'un long transcrit précurseur, processed dans un premier temps en *piARN* mature. Ce dernier pourra éventuellement entrer dans un processus d'amplification secondaire, la boucle d'amplification ping-pong, où les protéines Ago3 et Aub jouent un rôle essentiel (Brennecke et al., 2007; Gunawardane et al., 2007). Ces protéines interagissent mutuellement pour générer le pool de *piARN* secondaire, caractérisé par un biais en Adénine en position 10 de l'ARN. Cette signature « ping-pong » laissée sur les *piARN* secondaires résulte de l'activité de clivage du duplex Ago3/Aub, qui nécessite l'appariement de deux ARN complémentaires sur 10 paires de bases.

Dans la lignée germinale des animaux, les *piARN* sont essentiels au maintien de l'intégrité du génome. Dans la plupart des espèces étudiées à ce jour, la très grande majorité des *piARN* produits dérivent d'ET et sont activement impliqués dans leur répression transcriptionnelle et post-transcriptionnelle ((Siomi et al., 2011) pour une revue). Le rôle des *siARN* dans la répression des ET au sein de la lignée germinale est quant à lui, considéré comme mineure. Cependant, plusieurs travaux ont montré l'implication de *siARN* dans la répression post-transcriptionnelle des ET dans la lignée germinale. Chez le nématode (*Caenorhabditis elegans*), la plupart des mutants déficients de la voie des *siARN* sont associés à la dérégulation globale des ET (Ketjing et al., 1999; Tabara et al., 1999). Il a été proposé chez cette espèce que la transcription des éléments *Tc1* (super famille-*Tc1/mariner*) pouvait produire des ARN double brins complémentaires à partir des séquences terminales inversées répétées et par ce biais être réprimé entre *trans* par d'autres copies *via* des *siARN* (Sijen and Plasterk, 2003). Chez les mammifères, l'étude des allèles déficients de Dicer dans l'ovaire et l'ovocyte montre des niveaux plus élevés de transcription de certaines familles de rétrotransposons (Murchison et al., 2007; Svoboda, 2004). Par ailleurs, les séquençages de petits ARN dans l'ovaire, les ovocytes et les testicules de souris, montrent clairement que certains *siARN* dérivent de rétrotransposons de type *LINE* et ont la capacité de réprimer d'autres copies (Tam et al., 2008; Watanabe et al., 2006; Watanabe et al., 2008; Yang and Kazazian, 2006). De manière intéressante, une grande partie de ces *siARN* s'alignent sur des régions correspondant aux clusters de *piARN*, suggérant que ces « master loci » pourraient être impliqués à la fois dans la production de *siARN* et *piARN*. Il existe cependant certaines familles de rétrotransposons à LTR qui sont uniquement ciblées par des *siARN* (Tam et al., 2008; Watanabe et al., 2008). Chez *D. melanogaster*, le séquençage des petits ARN directement associés à la protéine Ago2 a révélé que les ET constituaient une source importante d'ARN d'une taille de 21 nt (Czech et al., 2008; Kawamura et al., 2008). Il a été montré par ailleurs, que leur accumulation dans la cellule était dépendante de la protéine Dcr2, qui est spécifiquement impliquée dans la production de *siARN* exogènes (Farazi et al., 2008). Enfin lorsque les protéines Ago2 et Dcr2 sont mutées, on observe systématiquement des niveaux plus élevés de transcrits dérivés d'ET (Chung et al., 2008; Czech et al., 2008; Kawamura et al., 2008). Ainsi, les voies de régulation par les *piARN* et les *siARN* sont toutes les deux activement impliquées dans la répression des ET. Certaines familles d'ET semblent être réprimées uniquement par l'une des deux voies, tandis que pour d'autres, il semble que ces deux mécanismes agissent de concert pour initier la production de petits ARN et réprimer transcriptionnellement les ET.

Ces observations concernent cependant des lignées d'ET qui partagent avec l'hôte une histoire évolutive établie sur le long terme. On ignore à ce jour la nature des interactions qui s'établissent entre le parasite et l'hôte lorsqu'une nouvelle famille d'ET émerge *de novo* dans un nouvel environnement génomique. L'émergence des ET dans un génome peut se produire *via* trois mécanismes différents (Capy, 1998; Loreto et al., 2008; Schaack et al., 2010). Premièrement, lorsque des génomes ne possédant pas l'élément, recombinent avec des génomes de la même espèce où l'élément est préexistant. Deuxièmement, par introgression, lorsque deux espèces proches s'hybrident (l'une possédant l'ET l'autre pas). Troisièmement, par transfert horizontal, notamment lorsque plusieurs espèces partagent un vecteur en commun (Ivancevic et al., 2013). La majorité des événements connus de transferts horizontaux entre espèces eucaryotes concerne des ET (Schaack et al., 2010). Cependant, les études focalisées sur de nouvelles invasions d'ET concernent en général des familles pour lesquelles les traces d'invasion passées sont toujours présentes dans le génome hôte. Ces transposons sont parfois encore présents dans leur forme canonique ou bien sous la forme de copies divergentes qui persistent dans les régions hétérochromatiques du génome (de Setta et al., 2009; Mejlumian et al., 2002).

L'élément *P* est l'un des ET les mieux caractérisés chez les eucaryotes. Il a été transféré horizontalement depuis l'espèce *Drosophila willistoni* (il y a environ 65 ans) chez *D. melanogaster*. L'élément *P* a été découvert initialement en tant que facteur provoquant l'émergence de phénotypes anormaux chez *D. melanogaster* (Kidwell and Novy, 1979). L'émergence de cet élément au sein de populations vides de l'élément s'est manifestée par l'apparition de recombinaisons chez les mâles, de taux élevés de stérilité et par

l'atrophie des gonades (syndrome de dysgenèse hybride). Le génome hôte s'est rapidement adapté à l'invasion de l'élément *P* via la production de *piARN* dérivés de l'élément, qui en répriment la transposition dans la lignée germinale. Il a été montré que les *piARN* produits pouvaient être transmis à la descendance de manière à initier la répression de l'élément *P* chez l'embryon et par ce biais le protéger des effets de dysgenèse hybride (Khurana et al., 2011). Par ailleurs, les mêmes auteurs ont montré que la baisse de fertilité induite ce syndrome pouvait être restauré au cours de la vie de l'individu par de nouvelles insertions de l'élément dans les clusters de *piARN*.

Le cas de l'élément *P* constitue l'un des rares exemples de transfert horizontal clairement démontré chez la drosophile. Il est cependant possible de « mimer » un transfert horizontal expérimentalement en introduisant l'élément par transgénèse et observer les moyens mis en œuvre par le génome pour limiter l'expansion de la nouvelle famille d'ET. Cette approche a été réalisée chez *D. melanogaster*, où l'élément *Penelope*, naturellement absent chez cette espèce a été introduit. L'élément *Penelope* transféré est celui de *Drosophila virilis*, séparée de *D. melanogaster* par 50 millions d'années de divergence évolutive. Chez *D. virilis*, l'élément est réprimé transcriptionnellement par les *siARN* dans la plupart des lignées sauvages (Blumenstiel and Hartl, 2005), à l'exception de la souche 160 où *Penelope* est réprimée par les *piARN* (Rozhkov et al., 2010a). De manière intéressante, après son introduction chez *D. melanogaster*, l'hôte a acquis la capacité à réprimer l'élément majoritairement par la production de *siARN* (Rozhkov et al., 2010a). Cependant, trois ans plus tard, ces mêmes souches ont été re-séquencées et il s'avère qu'aujourd'hui, l'élément *Penelope* est majoritairement réprimé par les *piARN* (Rozhkov et al., 2013). Ces résultats montrent la flexibilité des réponses apportées par le génome qui semblent varier dans le temps au cours du processus d'invasion de l'élément. Néanmoins, les travaux réalisés sur les éléments *P* et *Penelope* suggèrent que l'acquisition d'un processus de régulation stable chez *D. melanogaster* converge vers l'acquisition de copies productrices de *piARN*. Étendre ce type d'analyse à d'autres familles d'ET est fondamental pour comprendre dans quelle mesure l'invasion par de nouveaux parasites est une histoire répétable, indépendamment du caractère stochastique de chaque nouvelle insertion dans le génome.

Dans ce contexte, notre étude vise à comprendre l'établissement des interactions hôtes-parasites initiées par les voies de régulations des ARNi. L'élément *mos1* de la famille *mariner* (ET de Classe II), a été introduit chez *D. melanogaster* où il y est naturellement absent. Depuis son introduction (effectuée par Aurélie Hua-Van en 2000), l'élément a envahi le génome à hauteur de 20 copies et demeure stable depuis 10 ans. Il semble que l'élément soit entré dans une phase d'équilibre qui implique probablement l'intervention d'un processus régulateur. Dans ce travail, nous allons présenter les caractéristiques de l'activité de *mos1* suite à son invasion chez *D. melanogaster*. Par ailleurs, nous allons déterminer le mode de régulation auquel l'élément est soumis, dans quelle mesure la régulation affecte l'activité de *mos1* et dans quel contexte génomique la régulation a pu émerger.

Résultats

1 - La stabilité de l'élément *mos1* dans la lignée M19

L'élément transposable *mos1* est un transposon de Classe II qui se multiplie *via* un mécanisme de type « couper-coller ». Il est naturellement présent dans toutes les espèces du sous-groupe *melanogaster*, à l'exception de *D. melanogaster* qui a vraisemblablement perdu *mos1* il y a moins de 5 millions d'années (date de divergence entre *D. melanogaster* et son espèce sœur *D. simulans*). Les études menées sur l'élément dans les espèces du sous-groupe *melanogaster* ont montrées que *mos1* était toujours actif et régulièrement mobilisé. Cette mobilisation se manifeste par de fortes variations interspécifiques en terme de nombre de copies et par un polymorphisme d'insertions élevé (Brunet et al., 1999). Chez l'espèce cosmopolite *D. simulans*, l'activité d'excision (Picot et al., 2008), et le nombre de copies (Giraud and Capy, 1996; Woodruff et al., 1999) atteignent des différences remarquables entre populations. Cependant, on ignore à ce jour, les mécanismes à l'origine de ces variations. Afin de comprendre les relations qui s'établissent avec l'hôte au cours de l'invasion de *mos1*, ce dernier a été inséré par transgénèse dans la lignée *w¹¹¹⁸* en 2000 (Aurélien Hua-Van), avec comme vecteur un plasmide de type pCaSpeR (Figure 1A). La lignée transformée M19 initiale était caractérisée par l'insertion de *mos1* sur le chromosome II. L'élément *mos1* s'est par la suite excisé du site d'insertion initial puis s'est transposé de manière autonome au sein de la lignée M19 (Figure 1B). Le nombre de copies de l'élément a été quantifié par qPCR à différents temps et montre une remarquable stabilité depuis plusieurs années, autour de 20 copies par génome. En parallèle, le nombre de copies de *mos1* a été mesuré dans les deux sexes et ne montre aucune différence significative. Cela suggère les copies ne s'accumulent pas sur l'un des deux chromosomes sexuels, mais sont majoritairement distribuées sur les autosomes (Figure 1C). Les tests d'excisions de *mos1* en lignée somatique indiquaient que *mos1* était particulièrement actif (Picot et al., 2008). Nous montrons ici, que l'élément est bien transcrit chez les adultes des deux sexes (Figure 1D). Cependant, bien que le nombre de copies entre sexes soit identique, le niveau de transcription est trois fois supérieur chez les mâles que chez les femelles (Figure 1D).

Ces observations montrent donc que l'élément est toujours actif transcriptionnellement, bien que le nombre moyen de copies soit stabilisé et n'augmente plus. Cela suggère que le génome de la lignée M19 a probablement mis en place un mécanisme de régulation de l'activité de cet élément. Les différences entre mâles et femelles suggèrent par ailleurs, que ce mécanisme réprime l'expression de *mos1* spécifiquement chez les femelles.

2- L'expression de *mos1* est sensible aux déficiences des gènes de la voie des *piARN* uniquement

Deux mécanismes principaux ont été décrits pour leur rôle dans l'inhibition post-transcriptionnelle des ET dans la lignée germinale. Il s'agit de la voie des *piARN* et de la voie des *siARN* (Brennecke et al., 2007; Rozhkov et al., 2011). Nous avons tout d'abord caractérisé les activités transcriptionnelles de *mos1*, du gène *argonaute2* (*ago2*) associé à la régulation par les *siARN* ainsi que d'*argonaute3* (*ago3*) et *aubergine* (*aub*) associés à la régulation par la voie des *piARN* dans la lignée germinale (Figure Supplémentaire 1). De manière intéressante, on peut constater que les gènes associés à la régulation des ET présentent un patron d'expression biaisé dans les ovaires (Figure Supplémentaire 1A). Parallèlement, l'activité transcriptionnelle de *mos1* dans les testicules est environ 9 fois supérieure à celle retrouvée dans les ovaires de la lignée M19 (Figure Supplémentaire 1B). Par ailleurs, nous avons mesuré l'activité transcriptionnelle de *mos1* en contexte mutant pour différents gènes des deux voies de régulation (Figure 2). Le même schéma de croisement a été effectué pour l'ensemble des mutants (Cf Tableaux de croisements 1 à 5). Les données présentées Figure 2 montrent la comparaison entre mutants hétérozygotes et hétéroalléliques qui possèdent en commun le chromosome X, hérité de la mère M19 ayant l'élément *mos1*. Les mutants diffèrent au niveau des

chromosomes II et III : l'hétérozygote est maintenu *via* des chromosomes balancés, alors que chez l'hétéroallélique le fond génomique non balancé est issu des deux hétérozygotes. Les gènes *ago3* et *aub* sont associés à la régulation des ET par la voie des *piARN*. Ils interviennent plus précisément dans le processus de biogenèse secondaire associé à l'amplification massive des *piARN* à partir du répertoire de *piARN* issu du processus de biogenèse primaire. Lorsque l'on compare les niveaux d'expressions de *mos1* entre mutants hétérozygotes et hétéroalléliques, on observe une forte augmentation à la fois dans les testicules et dans les ovaires. Pour le gène *ago3*, l'expression de *mos1* est 12 fois supérieure dans les ovaires et 7 fois supérieure dans les testicules chez les mutants hétéroalléliques par rapport aux mutants hétérozygotes (Figure 2A). On observe un effet similaire pour les mutants hétéroalléliques du gène *aub*, pour lesquels l'expression de *mos1* est 6 fois plus importante dans les ovaires et 5 fois dans les testicules par rapport aux mutants hétérozygotes (Figure 2B). Par ailleurs, le gène *zucchini* (*zuc*) code pour une endonucléase indispensable au processus de biogenèse primaire des *piARN*. De même que pour les gènes *ago3* et *aub*, la comparaison entre mutants hétérozygotes et hétéroalléliques, montre que l'expression de *mos1* est 3 fois supérieure dans les ovaires du mutant hétéroallélique (Figure 2D).

La protéine codée par le gène *dicer2* (*dcr2*) forme un complexe avec la protéine R2D2 qui permet la prise en charge des *siARN* de manière à guider la reconnaissance des ARN messagers cibles et à initier leur dégradation. L'endonucléase Ago2 possède l'activité de clivage nécessaire à la dégradation des ARN messagers au sein du complexe RISC (Rand et al., 2005). La comparaison entre mutants hétérozygotes et homozygotes de l'allèle *dcr2* ne montre aucune différence significative quant à l'expression de *mos1*, suggérant que *dcr2* n'est pas impliqué dans la répression transcriptionnelle de l'élément, ni dans les testicules, ni dans les ovaires (Figure 2C). De la même manière, lorsque l'on compare, dans l'ovaire, l'expression de *mos1* entre les mutants hétérozygotes et hétéroalléliques du gène *ago2*, on n'observe aucune différence significative (Figure 2D).

L'ensemble de ces observations montre que l'activité transcriptionnelle de *mos1* est significativement accrue lorsque les gènes associés à la régulation par les *piARN* sont déficients. Au contraire, l'expression de l'élément semble indépendante des gènes associés à la régulation par les *siARN*. Il semble donc qu'au cours de son invasion par l'élément *mos1*, la lignée M19 ait acquis la capacité à réprimer transcriptionnellement l'élément, notamment *via* la production *de novo* de *piARN* dans la lignée germinale. Dans les croisements effectués, seul le chromosome X hérité de la mère M19 a été transmis. Dans le cas où les *piARN* seraient produits *de novo*, ces observations indiquent *i)* soit la présence d'une copie productrice de *piARN* sur le X (Josse et al., 2007) *ii)* soit la transmission épigénétique de *piARN* maternels capables d'initier la production de nouveaux *piARN* dérivés de *mos1* *en trans* (Grentzinger et al., 2012).

3- La production *de novo* de *piARN* dérivés de l'élément *mos1*

L'étude transcriptionnelle de *mos1* suggère donc que ce dernier est réprimé par la voie des *piARN* uniquement. Cela implique que de nouveaux petits ARN d'une taille spécifique (23-29 nt) et dérivés de l'élément *mos1* soient produits dans la lignée germinale. Pour confirmer cette hypothèse, nous avons effectué le séquençage des petits ARN dans les gonades mâles et femelles de la souche M19 (Figure 3). L'ensemble des petits ARN obtenus a été aligné sur la séquence de l'élément *mos1* et a été classé selon la taille dans un premier temps (Figures 3A et 3B). On observe la présence de petits ARN spécifiques à l'élément *mos1* tant dans les ovaires que dans les testicules de la lignée M19. Par ailleurs, aucun de ces petits ARN spécifiques de *mos1* ne s'aligne sur le génome de la souche *w¹¹¹⁸* (qui ne contient pas d'élément *mariner*). Inversement, aucun petits ARN issus de banques d'ovaires d'autres souches de *D. melanogaster* ne s'aligne sur la séquence de l'élément *mos1* (données non présentées). Ces observations montrent que suite à l'invasion de *mos1* dans le génome de la lignée M19, la production de nouveaux petits ARN a bien été initiée. Dans l'ovaire, la

distribution des petits ARN dérivés de *mos1* (Figure 3A), montre la présence majoritaire d'ARN sens et antisens de taille comprise entre 23 et 27 nucléotides avec un pic maximum de distribution situé à 25 et 26 nucléotides. De plus, ces petits ARN présentent une signature ping-pong significative. Ainsi, environ 15% des reads sens et antisens, présentent une Adénine en position 10 et un appariement sur 10 nucléotides avec un petit ARN complémentaire à partir de cette position. Ces observations montrent que l'élément *mos1* est capable de produire des *piARN* impliqués dans un processus de biogenèse secondaire. Dans les testicules, la majorité des petits ARN dérivé de *mos1* sont produits en sens et en antisens et leur taille est comprise entre 23 et 26 nucléotides avec un pic de distribution autour de 25 nucléotides (Figure 3B). De même que dans l'ovaire, environ 15% de ces petits ARN présentent une signature ping-pong significative (Figure 3B'). Néanmoins, les lignées germinales mâles et femelles diffèrent sur deux points: d'une part, la répartition des ARN sens et antisens et d'autre part, la quantité totale de *piARN* produits. En effet, la quantité totale de *piARN* (normalisée par million de petits ARN) produits dans les testicules est 10 fois inférieure à celle produite dans les ovaires (environ 2000 *piARN* chez les mâles contre 20 000 chez les femelles). On peut observer ces différences plus précisément *via* l'alignement des *piARN* (23-29 nt uniquement) sur la séquence canonique de l'élément *mos1* (Figures 3C et 3D). De manière surprenante, il semble que la production de *piARN* spécifiques de l'élément *mos1* soit nettement moins efficace dans la lignée germinale mâle. Ces résultats suggèrent donc des différences importantes de l'activité de la voie des *piARN* entre les deux lignées germinales. De plus, ces observations corroborent la surexpression transcriptionnelle de *mos1* dans les testicules et la sous expression des gènes *aub* et *ago3* dans ce même tissu (vérifiée au niveau protéique pour *ago3* – Figure Supplémentaire 2). Néanmoins, les pics de *piARN*, sens et antisens, le long de la séquence sont situés exactement aux mêmes positions (e.g. le pic de *piARN* antisens situé avant l'ITR 3'). Ce dernier point suggère que dans les deux lignées germinales, la production de *piARN* résulte d'un processus similaire probablement initié *i)* par une insertion dans un locus déjà producteur de *piARN* double brins (Brennecke et al., 2007), ou bien *ii)* par une nouvelle insertion capable d'initier *de novo* la formation d'un cluster de *piARN* (Shpiz et al., 2014).

4 – L'élément *mos1* transpose moins lorsqu'il est transmis maternellement

On sait que les *piARN* produits par la mère dans les cellules nourricières de l'ovaire peuvent être transmis à l'ovocyte de manière à activer la production de *piARN* secondaires chez l'embryon. Nous avons vu dans la section précédente que la lignée M19 avait acquis la capacité à produire de nouveaux *piARN* de manière à réprimer l'expression de *mos1*. On ignore cependant si ces *piARN* peuvent affecter la transposition de l'élément *mos1*. Pour répondre à cette question, nous avons effectué une série de rétrocroisements successifs entre la lignée M19 et des souches de *D. melanogaster* ne contenant pas l'élément *mos1* (Cf Tableau de croisement 6). Au cours de ces croisements, l'élément a été successivement apporté, soit par la mère, soit par le père sur 6 générations. Dans le cadre théorique où *mos1* ne transpose pas, on s'attend à ce que le nombre de copies soit diminué de moitié à chaque croisement. Au terme de ces 6 générations, le nombre moyen de copies attendu par individu est donc de 0.31 (en admettant que le nombre moyen de copies de la lignée M19 est de 20 copies par génome et que toutes soient sur les autosomes). Lorsque l'élément est systématiquement apporté par la mère, le nombre moyen de copies observé par individu est de 0.4 à la sixième génération, suggérant que l'élément ne transpose quasiment plus (Figure 4). En revanche, lorsque *mos1* est successivement transmis par les mâles, au terme des 6 générations, on compte en moyenne 3 copies par individus, suggérant que l'élément a été capable de se multiplier.

Alternativement, le fond génomique de la lignée M19 a été substitué en deux générations à l'aide d'une souche balancée sur les chromosomes I, II et III (Figure Supplémentaire 3 et Tableau de croisements 7). Ainsi le fond génomique de la lignée M19 a été complètement remplacé par le fond génomique de la souche balancée. La détection de *mos1*, au terme des deux générations, correspond donc à de nouveaux événements de transposition survenus en F1 ou en F2 chez l'embryon. Le génome de M19 a été substitué soit par la mère, soit

par le père et la présence de nouvelles insertions de *mos1* en F2 a été vérifiée sur des individus uniques par PCR. Lorsque que l'on compare le pourcentage d'individus contenant de nouvelles insertions, moins de 60% des individus possèdent de nouvelles insertions de *mos1* lorsque le fond génomique est éliminé par la mère contre 90% lorsque celui-ci est substitué par le père. Ces résultats suggèrent que chez l'hybride maternel (M19/souche balancée), les *piARN* transmis maternellement ont limité la transposition de l'élément *mos1*. En parallèle, l'absence de copies sur le Y de la lignée M19 a été confirmée en introduisant ce dernier dans nouveau fond génomique (Cf Tableau de croisements 8).

5- Un équilibre dynamique

La lignée M19 est capable d'initier la production de *piARN* spécifiques de l'élément *mos1*, ce qui suggère qu'au moins une insertion soit capable d'en assurer la production. Cette lignée compte en moyenne 20 copies par génome depuis plusieurs années et la localisation des nouvelles insertions pourrait être informative sur la nature des copies à l'origine de la production *de novo* de *piARN*. Trois méthodes différentes ont été employées dans le but de localiser les insertions de l'élément *mos1*. A l'aide d'une sonde construite sur toute la longueur de l'élément nous avons pu identifier sur les chromosomes polytènes de deux individus, un total de 39 régions cytologiques différentes (Tableau 1). Cependant, seulement 8 insertions sont communes aux deux individus, 23 semblent spécifiques à un seul des deux individus (*i.e.* retrouvées dans plusieurs noyaux chez un seul individu) et 8 autres n'ont été identifiées qu'à une reprise dans un seul noyau. De manière surprenante, on peut noter que, malgré l'apparente stabilité en termes de nombre de copies, ces dernières occupent majoritairement des localisations distinctes, suggérant la mobilisation récente de *mos1*.

En parallèle, à partir de l'ADN génomique de la lignée M19 (Cf méthodes), 71 séquences contenant une copie *mos1* et ainsi que les régions adjacentes de l'élément ont pu être identifiées. Les séquences adjacentes ont ensuite été *blastées* sur le génome de référence de manière à les localiser (Tableau 2 et 3). Les deux tableaux font la distinction entre les séquences que l'on peut localiser de manière exacte dans le génome (Tableau 2) et celles qui s'alignent sur des régions répétées du génome (Tableau 3). Parmi les 47 séquences ayant une localisation unique, nous avons pu identifier 24 insertions indépendantes de *mos1* au niveau d'introns de gènes. Parmi les 26 séquences ayant une localisation multiples, toutes correspondent à des insertions au niveau d'ET. Malgré la répression par les *piARN*, ces observations montrent que le polymorphisme d'insertion est très important, indiquant que l'élément *mos1* a été récemment actif et continue vraisemblablement de transposer.

Il a récemment été suggéré qu'un moyen efficace de réprimer l'insertion de nouveaux ET viendrait de leur capacité à générer de nouveaux clusters de *piARN* simplement par leur insertion *de novo* dans l'euchromatine (Shpiz et al., 2014). L'hypothèse alternative est que les régions associées à la régulation par les *piARN* constituent un nombre limité de *loci* génomiques, conservés au cours de l'évolution (Malone and Hannon, 2009) et que la prise en charge d'un nouvel ET dans le génome passe nécessairement par son insertion dans ces régions spécifiquement dédiées à la production de *piARN*. Parmi les deux types d'insertions identifiées ici, (*i.e.* localisation génomique unique ou multiple), toutes les régions flanquantes de l'élément ont été extraites et les *piARN* issus des ovaires de M19 y ont été alignés. Il est intéressant de constater que les régions flanquantes répétées contiennent en moyenne 4,2 *piARN* par nucléotide (Figure 5A), ce qui est presque trois fois supérieur aux régions flanquantes à localisation unique (1.5 *piARN* par nucléotide – Figure 5B). En outre, environ 15% des *piARN* issus de régions répétées possèdent une signature ping-pong contre seulement 5% pour ceux issus de régions uniques (le détail par séquence est donné Figure Supplémentaire 4). Plus précisément, pour les régions flanquantes dont la localisation est unique, la très grande majorité des *piARN* qui s'y aligne, provient de deux insertions indépendantes qui contiennent les éléments transposables G2 et 297 (seqB et seqE du Tableau 2). Lorsque ces deux séquences sont retirées, on n'observe plus que 0,08

piARN par nucléotide dans ces régions. Ces résultats montrent que les insertions euchromatiques de *mos1* sont incapables d'initier la production de *piARN* dans les régions avoisinantes, comme il avait été suggéré par Shpiz (Shpiz et al., 2014). Au contraire, la très grande majorité des *piARN* dérivés des séquences flanquantes provient de régions répétées. Ces observations indiquent que pour achever sa régulation, l'élément *mos1* a été inséré dans une région riche en ET déjà productrice de *piARN*.

Discussion

Les ET évoluent rapidement et leur mobilisation peut déclencher des instabilités génétiques majeures, comme par exemple, les syndromes de dysgenèse hybride observés chez plusieurs espèces de drosophiles (Blumenstiel and Hartl, 2005; Bucheton et al., 1984; Kidwell and Novy, 1979). La voie des *piARN* a un rôle qui semble conservé dans le processus de répression des ET, mais les mécanismes par lesquels ce système s'adapte à l'émergence de nouveaux ET dans le génome sont peu étudiés. Par l'utilisation de la souche M19, dans laquelle l'élément *mos1* a été introduit, nous avons montré que l'élément s'est dans un premier temps amplifié de manière autonome et envahi le génome à hauteur de 20 copies. Depuis 10 années au moins, le nombre de copies est demeuré stable suggérant l'intervention de mécanismes impliqués dans la répression de l'activité de *mos1*. Par l'étude de l'activité transcriptionnelle de *mos1* en contextes mutants, nous avons pu montrer que les gènes associés à la régulation par les *piARN* étaient impliqués dans la répression de l'élément. Le séquençage des petits ARN a révélé que des *piARN* (23-29 nt) dérivés de *mos1* étaient produits *de novo* dans les lignées germinales mâles et femelles. L'étude des *piARN* sens et antisens avec une signature ping-pong, indique que ceux-ci sont produits au niveau de cluster de *piARN* par un processus d'amplification secondaire. Nous avons montré en parallèle que la transmission maternelle de l'élément est associée à une activité moindre de *mos1*, suggérant que les *piARN* produits sont transmis par la mère et répriment partiellement la transposition de *mos1*. Par ailleurs, 71 séquences contenant une copie de *mos1* et les régions flanquantes ont été identifiées. Les *piARN* des ovaires de la lignée M19 ont été alignés sur ces régions flanquantes et seules les régions apparentées à des ET sont associées à un enrichissement en *piARN*, indiquant que la ou les copies régulatrices de *mos1* sont situées dans des régions hautement répétées semblables à des clusters de *piARN*.

La voie des *piARN* limite l'expansion des nouveaux ET

Suite à l'invasion de *mos1* dans la lignée M19, le génome a acquis la capacité à réprimer transcriptionnellement l'élément par la voie des *piARN*. L'exemple de *mos1* n'est pas unique puisque, suite à leur introduction chez *D. melanogaster*, d'autres familles d'ET montrent que la machinerie des *piARN* prend en charge ses nouveaux arrivants. L'élément *P* et l'élément *I* ont été transférés horizontalement chez *D. melanogaster* et, tous deux sont aujourd'hui activement réprimés cette voie (Kelleher et al., 2012; Khurana et al., 2011). Ceci dit, la mise en place de cette régulation n'est pas toujours la même. Ainsi, l'élément *Penelope*, a été introduit chez cette même espèce et dans les premiers temps de l'invasion le génome produisait des *siARN* ciblant spécifiquement l'élément (Rozhkov et al., 2010b). Cependant, trois années plus tard, le séquençage des mêmes souches a révélé que *Penelope* était activement réprimé par la voie des *piARN* (Rozhkov et al., 2013). L'ensemble de ces observations montre qu'indépendamment du type d'élément nouvellement introduit le génome, *D. melanogaster* possède la capacité de s'adapter à l'invasion de nouveaux parasites par la mise à jour son répertoire de *piARN*. Cependant, on peut se demander dans quelle mesure ce phénomène est systématique. En effet, bien que la production *de novo* de *piARN* spécifiques ait été détectée suite à l'introduction de *Penelope* chez *D. melanogaster*, une partie seulement des lignées transformées ont acquis cette capacité. De la même manière, chez *D. virilis*, l'étude des petits ARN dérivés de l'élément *Penelope* chez plusieurs lignées montre une hétérogénéité de réponses. Selon les lignées, la répression de l'élément est initiée, soit par les *siARN*, soit par les *piARN*. Néanmoins, parmi ces lignées, toutes celles contenant des copies complètes de *Penelope* sont productrices de *piARN* dérivés de l'élément (bien que ces populations soient issues de localisations géographiques distinctes). Par ailleurs, au vu d'analyses plus globales, il ne semble pas que les nouveaux ET soient spécifiquement ciblés par la machinerie des *piARN*. Une étude s'intéressant aux populations de *piARN* présents dans les ovaires de *D. melanogaster* montre que

les transposons les plus jeunes (ceux qui transposent encore activement mais ne sont pas fixés dans le génome) sont très peu ciblés par les *piARN* (Kelleher and Barbash, 2013). Au contraire, il semble que la majorité des *piARN* ciblent des transposons qui ont une origine ancienne et pour lesquels on retrouve des copies fragmentées dans les clusters de *piARN* (Kelleher and Barbash, 2013). L'ensemble de ces résultats indiquent d'une part, que la réponse apportée par les *piARN* comporte un caractère stochastique (vraisemblablement liée à la fréquence d'insertion de nouvelles copies dans les clusters de *piARN*) et d'autre part, que cette réponse n'est pas immédiate (puisque les jeunes ET actifs sont globalement minoritaire dans les populations de *piARN*).

La dynamique de *mos1* dans le génome : un équilibre instable

L'invasion d'un ET dans une population de génome comprend trois phases distinctes (Blumenstiel and Hartl, 2005). Premièrement une phase exponentielle de croissance où l'élément transpose activement et s'établit dans la population, deuxièmement une phase stationnaire où le nombre copies demeure stable (équilibre entre gain et perte de copies), troisièmement une phase de décroissance associée à la perte aléatoire des copies en cours de dégénérescence. Nos résultats suggèrent que l'élément *mos1* est dans la phase stationnaire de la dynamique d'invasion puisque le nombre de copies semble stable depuis au moins 10 ans. Nous avons localisé les insertions de *mos1* sur les chromosomes de la lignée M19 par hybridation *in situ* (Tableau 1) et on compte environ 20 copies par individu, ce qui corrobore les données relatives à la quantification de l'élément par qPCR (Figure 1). Deux individus ont été analysés et comptent en commun, huit insertions localisables au niveau de bandes cytologiques identiques. Plusieurs de ces insertions ont été retrouvées et identifiées par PCR séquençage sur des pools d'individus indépendants, suggérant que certaines copies sont établies de manière stable dans la lignée. Par exemple, la copie localisée sur la bande cytologique 63B chez les deux individus, a été retrouvée dans un pool d'individus indépendant extrait en 2015. Par ailleurs, cette insertion était déjà présente au sein d'un pool d'individus extrait en 2010. De la même manière la bande 43E, localisée *in situ*, a pu être séquencée à partir d'extractions datées de 2015 et 2008. Néanmoins, malgré cette apparente stabilité pour certaines copies, nous avons constaté un polymorphisme d'insertions assez élevé, indiquant que l'élément *mos1* a été mobilisé récemment (Tableau 1). Du point de vue de la dynamique de l'élément, cela suggère que les nouvelles insertions de *mos1* sont régulièrement compensées par la perte d'autres. Nous avons vu par ailleurs que *mos1* était transcriptionnellement réprimé par la voie des *piARN* (Figure 2) et que cela affectait sa capacité à transposer (Figure 4). Notre hypothèse est que le taux de transposition de *mos1* est limité par l'intervention de la machinerie des *piARN*, de telle sorte que le nombre de nouvelles insertions soit en moyenne égal au nombre de délétions de l'élément.

De manière générale la production de *piARN* dépend de la production de *piARN* primaires qui agissent comme des initiateurs de la boucle d'amplification secondaire (Brennecke et al., 2007; Handler et al., 2011), comme c'est le cas pour une majorité d'éléments (e.g. *RT1b* chez *D. melanogaster*). Cependant, pour d'autres ET, comme les éléments *F* et *Doc*, la production de *piARN* secondaires est bien moins affectée par des mutations de gènes impliqués dans la biogenèse primaire, indiquant que l'initiation du système dépend essentiellement de *piARN* transmis maternellement qui agissent comme des initiateurs (Handler et al., 2011; Malone and Hannon, 2009). Ce type d'effet maternel a été formellement démontré par plusieurs études. Pour l'élément *P*, il a été montré que, les *piARN* massivement produits (pris en charge par les protéines de la sous famille PIWI) au niveau des cellules nourricières de l'ovaire, pouvaient être transmis à l'ovocyte et par ce biais activer la répression de l'élément, d'une part chez l'embryon (Brennecke et al., 2008) et d'autre part à travers les générations (de Vanssay et al., 2012). Pour l'élément *I*, l'absence de *piARN* secondaires hérités par la mère entraîne l'absence de production de *piARN* secondaires dans la descendance. Dans notre cas, les données suggèrent que l'expansion de *mos1* est limitée par la transmission épigénétique de petits ARN dérivés de l'élément. En effet, nous avons montré, *via* plusieurs croisements indépendants (Figure 4, Figure

Supplémentaire 3) que la transposition de *mos1* était réduite lorsque celui-ci était transmis maternellement. Par ailleurs, l'activité transcriptionnelle de *mos1* dans l'ovaire semble moins affectée en contexte mutant pour le gène *zuc* (gène impliqué dans le processus de biogenèse primaire), comparé aux gènes *ago3* et *aub* (Figure 2). Cet effet est plus marqué dans les testicules où on n'observe aucune différence entre les mutants hétérozygotes et hétéroalléliques du gène *zuc* (données non présentées). Ces résultats indiquent que l'élément est surtout sensible à une déficience de la voie de biogenèse secondaire des *piARN*. Il est cependant nécessaire de confirmer cette hypothèse par le séquençage des *piARN* dans les ovaires de deux hybrides issus d'un croisement entre la lignée M19 et une souche vide de l'élément (e.g. *w¹¹¹⁸*).

L'ensemble de ces observations suggère que la phase d'équilibre associée à la dynamique de *mos1* dans la lignée M19 est probablement la conséquence de *piARN* hérités maternellement qui déclenchent sa répression transcriptionnelle à travers les générations. La production de *piARN* dérivés de *mos1* implique vraisemblablement la présence d'au moins une copie productrice capable de réprimer en *trans* l'expression des autres copies. Les moyens par lesquels la machinerie des *piARN* s'adapte à l'arrivée de nouveaux transposons sont encore mal connus. Dans le but de comprendre l'environnement génomique favorable à l'émergence d'une régulation *de novo*, nous avons localisé un certain nombre de copies et caractérisé les régions flanquantes à ces insertions.

Deux modèles d'adaptation à l'émergence de nouveaux ET dans le génome hôte :

L'apparition *de novo* de clusters de *piARN* vs la sélection de clusters pré-existant

Il existe dans la littérature deux modèles permettant d'expliquer l'acquisition d'un système de répression des nouveaux ET par la voie des *piARN*. Le premier est basé sur l'observation que la production de *piARN* est majoritairement initiée à partir d'un petit nombre de *loci* dans le génome, appelés les « master loci » (Brennecke et al., 2007). La localisation de ces « master loci » est conservée chez plusieurs espèces mais le contenu en ET dans ces régions y est extrêmement variable (Malone and Hannon, 2009). Il a donc été proposé que la production *de novo* de *piARN* permettant de réprimer un nouvel ET nécessite une insertion de ce dernier dans un cluster de *piARN* pré-existant. Le second modèle est basé sur l'observation que de nouvelles insertions euchromatiques peuvent générer spontanément de nouveaux *piARN*, dont la production s'étend aux régions flanquantes à cette nouvelle insertion (Olovnikov et al., 2013; Shpiz et al., 2014). L'hypothèse supportée par ces auteurs est qu'une telle insertion est capable de recruter le matériel protéique nécessaire à la production de *piARN* et qu'à terme, cela donne naissance à de nouveaux clusters de *piARN*. Une des limites de ce modèle est que l'émergence de nouveaux clusters nécessite qu'en amont, l'élément soit déjà réprimé par les *piARN*. Plus précisément, ce modèle repose sur la capacité des *piARN* (produits à un locus déjà existant) à guider (*via* l'ARN guide) la machinerie de production des *piARN* à un autre locus.

Dans notre étude nous avons pu identifier les régions flanquantes de 71 insertions de *mos1* et déterminer leur localisation génomique. Parmi ces 71 insertions, 47 sont localisables au sein de 31 *loci* uniques dans le génome de *D. melanogaster* (Tableau 2), tandis que les 26 autres sont localisables au sein de régions répétées. Parmi les 31 localisations uniques, 24 correspondent à des insertions euchromatiques situées dans des introns de gènes et 7 sont situées dans des régions intergéniques. Par ailleurs, les 26 insertions à localisation multiples correspondent toutes à des insertions au sein de 12 familles d'ET différentes. Ici, nous avons donc deux types d'insertions (régions uniques vs régions répétées) pour lesquelles nous avons testé la présence de *piARN* ou non (Figure 5A et B). De manière intéressante, nous avons observé un fort enrichissement en *piARN* dans les régions apparentées aux ET (Figure 5A) alors que ce n'est pas le cas dans les régions non répétées du génome (Figure 5B). On constate un enrichissement similaires en *piARN* issus de la boucle d'amplification ping-pong pour les insertions de *mos1* présentes dans les ET (Figure 5A). Par ailleurs, la quasi totalité des petits ARN dérivés des flanquantes à localisation unique, provient de trois séquences contenant les

éléments G2 et 297. Lorsque ces séquences sont retirées du jeu de données, on ne compte plus que 0.08 *piARN* par nucléotide (Figure 5C). Ces observations ne supportent pas le modèle selon lequel de nouvelles insertions euchromatiques puissent initier *de novo* la formation de cluster de *piARN*. Il semble au contraire que seules les insertions de *mos1* situées dans des ET soient capables de produire de nouveaux *piARN*. Cela indique que les *piARN* dérivés de *mos1* résultent vraisemblablement d'une ou de plusieurs insertions dans des clusters de *piARN* pré-existant, caractérisés par leur forte densité en ET tronqués. Par ailleurs, du point de vue évolutif, générer de nouveaux clusters dans des introns de gènes n'est probablement pas avantageux, surtout lorsque ces gènes sont essentiels pour l'hôte. Une étude parue récemment soutient cette hypothèse (Lee, 2015) et montre que de manière générale, une insertion d'ET située à proximité d'un gène, ségrège à basse fréquence dans les populations. De manière intéressante, cette idée repose sur les mêmes principes que ceux favorable à l'émergence de nouveaux clusters de *piARN* via une insertion euchromatique. Cela suppose qu'un ET ciblé par les *piARN* puisse 1) recruter localement les protéines impliquées dans la méthylation des histones et 2) permettre par ce biais l'hétérochromatinisation des régions flanquantes à l'insertion. Dans le cas où une insertion de ce type arrive au sein d'un gène, la transcription de ce dernier peut être fortement réduite. Dans notre cas, nous avons détecté une insertion de *mos1* (seqT du Tableau 2) dans le gène *cnt1* dont la mutation est létale et une autre dans le gène *fs(1)h*, aussi associé à fort taux de mortalité (seqV du Tableau 2). Ces insertions n'ont été détecté qu'une fois, suggérant qu'elles sont rares dans la lignée et vraisemblablement non retenues par la sélection naturelle.

L'analyse des petits ARN, mappant sur les 3 régions uniques du génome contenant des fragments d'ET (séquence B, D et E du Tableau 2), a révélé certaines propriétés singulières. En effet, le pic majoritaire des petits ARN correspond à une taille de 21 nucléotides, la taille caractéristique des *siARN* (Figure 5B). A ces *loci*, il semble qu'il y ait production, à la fois de *piARN* et de *siARN*. Chez le nématode *C. elegans*, la production de *piARN* (21 nt chez cette espèce) peut déclencher la biogenèse secondaire de *siARN*, majoritairement impliqués dans la défense contre les éléments mobiles (Bagijn et al., 2012). Chez la drosophile, plusieurs cas de nouvelles insertions associées à la production de *siARN* endogènes ont été décrits (Olovnikov et al., 2013; Shpiz et al., 2014). Il a été montré par ailleurs que *siARN* et *piARN* étaient capables de s'associer en paires et générer une signature ping-pong (Shpiz et al., 2014), suggérant la possibilité que la voie des *piARN* et *siARN* puissent coopérer dans la lignée germinale pour amplifier les petits ARN depuis une nouvelle insertion d'ET. Néanmoins, ces rares cas sont *a priori* liés au contexte génomique dans lequel l'insertion s'est produite (Olovnikov et al., 2013). Par exemple, dans notre cas il s'agit d'une insertion dans un solo LTR au milieu d'un intron (e.g. seqD Tableau 2).

***mos1* est différentiellement régulé entre lignées germinales mâles et femelles**

La transcription de l'élément *mos1* en condition sauvage pour la lignée M19 montre des différences importantes entre les deux lignées germinales. En effet, l'expression de *mos1* dans les testicules est 8 fois supérieure aux niveaux d'expressions détectés dans l'ovaire. En revanche, c'est le contraire pour les gènes associés à la régulation par ARNi : ceux-ci sont systématiquement surexprimés dans l'ovaire comparé aux testicules. Ces observations corroborent par ailleurs la faible quantité de *piARN* dérivés de *mos1* produite dans les testicules en comparaison aux ovaires. De plus, selon le mutant et le sexe de la lignée germinale étudiés, les niveaux de transcription de *mos1* comportent des différences. Nos observations suggèrent que la lignée germinale mâle est plus sensible à une déficience du gène *ago3*, qu'à une déficience du gène *aub* (Figure 2). En effet, lorsque l'on compare les niveaux d'expressions de *mos1* entre mâles et femelles des lignées hétéroalléliques, *mos1* est deux fois plus exprimé chez les mâles lorsqu'*ago3* est muté et non différentiellement exprimé lorsqu'*aub* est muté. De manière intéressante, ces observations sont en accord avec le séquençage des *piARN* spécifiquement liés aux protéines Ago3 et Aub dans les testicules (Nagao et al., 2010). Dans cette étude il est montré que 53% des *piARN* dérivés d'ET sont associés à Ago3 contre seulement

7% pour Aub. Par ailleurs, Nagao *et al.* montrent que la protéine Ago3 localise uniquement dans une petite population de cellules au niveau de l'apex du testicule. Cette région correspond aux cellules indifférenciées de la lignée germinale où la machinerie des *piARN* a démontré son rôle pour le maintien du renouvellement des souches (Juliano *et al.*, 2011; Klenov *et al.*, 2011). L'ensemble de ces résultats indique que dans les testicules, la machinerie des *piARN* n'est effective que dans le pool de cellules souches, bien avant que le processus de gamétogenèse n'ait été initié. La localisation des transcrits de *mos1* par hybridation *in situ* corrobore cette hypothèse et indique que l'élément est transcriptionnellement actif dans les spermatocytes, c'est-à-dire après la différenciation des cellules souches (Figure Supplémentaire 5). L'ensemble de ces observations montre clairement des différences majeures entre les deux lignées germinales. La répression de *mos1* (et potentiellement d'autres ET) dans les testicules semble limitée et semble absente dans les stades correspondant à la formation des gamètes. L'absence de régulation de *mos1* à ces stades peut en favoriser la transmission à travers les générations. Dans les testicules, *mos1* participe vraisemblablement au fort polymorphisme d'insertions dans la lignée M19 qui indique la mobilisation récente de l'élément. La mobilisation de *mos1* dans la lignée germinale mâle, contre balancée par sa forte répression dans l'ovaire, fait échos au paradoxe de l'équilibre instable, où le nombre de copies demeure en moyenne constant à travers les générations bien que l'élément continue d'être mobilisé.

Matériel et méthodes

Souches et croisements

La souche M19, a été construite à partir de la lignée w^{1118} , dans laquelle l'élément transposable *mosI* a été introduit par transgénèse en 2000 (Aurélien Hua-van). Un plasmide de type pCaSpeR a été utilisé comme vecteur, contenant la séquence complète de l'élément *mosI* et un marqueur *mini-white*, avant d'être injecté par une méthode standard (Rubin and Spradling, 1982). Le génotype de la souche M19 initiale était: Xw^{1118}/Y ; $P\{mosI, w^+\}/P\{mosI, w^+\}$; III/III ; IV/IV.

A partir de la souche M19, une série de croisements a été réalisée avec des souches déficientes pour des gènes de la voie des *piARN* et des *siARN*. *Zucchini* (*zuc*) et *aubergine* (*aub*) sont situés sur le chromosome 2 et maintenus par un balancier *Cyo*. *Argonaute 3* (*ago3*) situé sur le chromosome 3, est maintenu par un balancier *TM6B*. La perte de fonction de ces gènes engendre une descendance stérile. Les allèles *zuc* ont été isolés à la suite d'un criblage EMS (Schupbach and Wieschaus, 1991). *zuc^{HM27}* contient un codon stop au niveau du résidu 5 et *zuc^{SG63}* est substituée par une Tyrosine au niveau de l'Histidine 169 (domaine HKD impliqué dans l'activité nucléase de la protéine). L'allèle *aub^{QC42}* a été isolé par crible EMS (Schupbach and Wieschaus, 1991) et l'allèle *aub^{N11}* contient une délétion de 154 paires de bases induisant un décalage du cadre de lecture (Wilson et al., 1996). Les allèles *ago3^{t2}* et *ago3^{t3}* ont été obtenus à partir d'un crible EMS (Koundakjian et al., 2004) et contiennent tout les deux un codon stop prématuré au niveau de leur séquence (Hain et al., 2010). Les allèles *argonaute 2* (*ago2*) sont situés sur le chromosome 3 et maintenus par un balancier *TM3*, *Sb[1]* et *TM6B*, *Tb[1]* respectivement pour les allèles *ago2³²¹* et *ago2^{dop1}*. L'allèle *ago2^{dop1}* contient une délétion de 69 nucléotides (en phase) au niveau du domaine GRR de l'exon 3 (Meyer et al., 2006); l'allèle *ago2³²¹* contient une délétion de 2514 paires de bases (décrite dans (Hain et al., 2010). Les embryons issus d'un croisement *ago³²¹/ago^{dop1}*, donnent une descendance viable (Hain et al., 2010). L'allèle *Dicer-2* (*dcr-2^{L811fsX}*) a été obtenu par criblage EMS (Lee and Carthew, 2003), contient 6 substitutions (Lee et al., 2004) et il est maintenu par un balancier *Cyo*. Les mutants homozygotes pour cet allèle donnent une descendance viable. Les souches ont été fournies généreusement par Stéphane Ronsseray.

Quantification du nombre de copies

L'ADN génomique a été extrait au Chelex 100 (Biorad) selon les recommandations du fabricant. Les estimations du nombre de copies de *mosI* ont été effectuées par RT-qPCR à partir de l'ADN génomique. Le nombre de cycles d'amplification (CT) obtenu pour *mosI* a été normalisé par la souche *WP5S* qui ne contient qu'une seule copie de l'élément. Les estimations du nombre de copies représentent la moyenne de trois répliques biologiques effectués en parallèle sur un pool d'individus de la même génération. La méthode des $\Delta\Delta Cq$ a été utilisée avec *RPL17q2* comme gène de référence. Les primers utilisés ont été définis dans les régions conservées sur l'ensemble des copies connues appartenant à la sous-famille *mosI*.

Quantification des niveaux transcriptionnels d'ARN

Entre 30 et 50 individus ont été extraits pour chaque mesure expérimentale. Les mesures présentées sont la moyenne de trois répliques biologiques. Les tissus ont été disséqués dans du PBS et conservés dans un tampon RA1 (Macherey-Nagel) et stockés à -80°C. Les échantillons ont été broyés par des billes de métal (3mm) avant l'extraction des ARN totaux. Les extractions d'ARN totaux ont été effectuées avec le kit NucleoSpin®

RNA, selon les recommandations du fabricant. Les concentrations d'ARN ont été quantifiées par absorbance à 260nm sur un spectrophotomètre NanoDrop 2000. La qualité des ARN totaux a été estimée à l'aide du système Experion™ Automated Electrophoresis System (Bio-Rad). Seuls les échantillons de haute qualité (RQI > 7) ont été conservés à -80°C. De manière à éviter les contaminations par de l'ADN génomique, une deuxième digestion d'ADN a été effectuée sur ces échantillons, selon les recommandations du constructeur (Turbo DNA-free™ - Ambion). Les synthèses d'ADN complémentaires à partir des ARN totaux (concentration 1µg/µl) ont été réalisées à l'aide du kit iScript™ cDNA synthesis (Bio-Rad) et conservés à -20°C.

L'expression des gènes a été quantifiée *via* le système de détection CFX96™ Real-Time PCR (Bio-Rad), et à l'aide de l'iQ SYBR® Green Supermix (Bio-Rad). L'expression des gènes a été normalisée sur la base de plusieurs gènes de références et par la méthode des $\Delta\Delta Cq$. Pour chaque condition expérimentale, les gènes de ménage les plus stables ont été sélectionnés par trois logiciels de statistiques dédiés : BestKeeper (Pfaffl et al., 2004) ; NormFinder (Andersen et al., 2004) ; geNorm (Vandesompele et al., 2002).

Hybridation *in situ* sur transcrits

Les expériences d'ARN FISH ont été réalisées de manière à localiser les transcrits de l'élément *mos1* dans les testicules. Cinquante paires de testicules ont été disséquées dans un tampon (10mM Tris-HCl, 47mM NaCl, 183mM KCl) et les expériences d'ARN FISH menées selon les procédures décrites par (Morris et al., 2009). La sonde a été synthétisée avec le kit TA cloning (Invitrogen) et les amorces définies selon le protocole du fabricant.

Western Blot

Nous avons quantifié les niveaux d'expression protéiques d'Ago3 par western blot. Cinquante paires de testicules et 30 paires d'ovaires ont été lysés dans 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Triton-X100, 5% glycérol, pH 7.5. Les niveaux de protéines ont été quantifiés par Bradford (Bradford assay – Biorad Protein Assay), et des quantités identiques de protéines ont été chargées sur gel (35µg). Les SDS-PAGE ont été réalisés en utilisant des gels NuPAGE 4-12 % Bis-Tris (Life Technologies) dans du tampon MES (MES 50mM, Tris-Base 50mM, SDS 0.1 %, EDTA 1mM, pH 7.3). Les protéines ont été transférées en utilisant le system I-blot (Life Technologies). L'anticorps Ago3 a été fourni par J. Brennecke (2007). Les dilutions de l'anticorps ont été effectuées au 2000^{ème}. Nous avons utilisé des anticorps couplés AP (Promega) pour la révélation protéique.

Extraction et séquençage des petits ARN

L'extraction totale d'ARN a été effectuée à partir de 50 paires et 10 paires d'ovaires et de testicules respectivement. Les dissections ont été effectuées de 2 à 4 jours après émergence des pupes. Un protocole d'extraction au trizol a été utilisé selon les procédures du fabricant (TRIzol Reagent, Invitrogène). Les petits ARN ont été sélectionnés pour leur taille à partir d'1µg d'ARN total sur un gel d'acrylamide 15 %. Les ARN résultants ont été traités à l'aide du kit Illumina « Truseq Small RNA sample prep » selon les recommandations du fabricant. Les bibliothèques ont été séquencées *via* l'appareil Illumina HiSeq 1000 instrument, et par l'utilisation des kit TruSeq SR Cluster v3-cBot-HS (Illumina) et TruSeq SBS v3-HS – 50 cycles (Illumina), selon le protocole adapté au séquençage de lectures simple brin d'une longueur de 50 paires de bases. Les bibliothèques ont été rassemblées dans des proportions équimolaires et diluées à une concentration finale de 12pM, selon les

recommandations d'Illumina. Les données ont ensuite été démultiplexées à l'aide du logiciel CASAVA (CASAVA_1.8.2). La qualité des données a été vérifiée à l'aide du logiciel FastQC 0.10.1.

Localisation des copies génomiques

a) Détection à partir des *piARN* flanquants l'élément *mos1*

A partir des bibliothèques de *piARN* d'ovaires de la lignée M19, nous avons isolé les séquences flanquantes de part et d'autre de plusieurs sites d'insertions de l'élément *mos1*. Toutes les combinaisons de séquences flanquantes des extrémités 5' et 3' de l'élément ont été reconstruites, puis blastées dans le génome de référence, dans le but de localiser les sites d'insertions. L'ensemble des régions du génome s'alignant avec les sites d'insertions potentiels a été importé (2kb de part et d'autre du site d'insertion). Dans le but de sélectionner uniquement les *loci* riches en *piARN*, les banques de *piARN* d'ovaires ont été alignées sur ces régions de 4kb. Seules les régions s'alignant avec au moins 4000 *piARN* (1 *piARN* par position en moyenne) ont été conservées, puis testées individuellement par PCR pour vérifier la présence de l'élément *mos1*. Pour se faire, les PCR ont été réalisées avec une amorce fixe à l'intérieur de l'élément et une amorce spécifique dans la région flanquante potentielle. Les insertions ont été discriminées sur gel d'agarose (par la présence d'une bande et sa taille), puis vérifiées individuellement par séquençage.

b) Détection des insertions par splinkerette

L'ADN génomique a été extrait à partir de 40 individus de la lignée M19 à l'aide du kit Nucleospin® Tissue (Macherey Nagel), selon les recommandations du fabricant. L'ADN obtenu (80ng) a été digéré par l'enzyme Fast Digest (ThermoFisher), selon les recommandations. Une fois digéré, une étape de ligation (Ligase 400U/μl) est effectuée en présence des adaptateurs splinkerette (15ng/μl) dans le « Fast Digest Green Buffer » du kit Fast Digest (ThermoFisher) toujours selon les recommandations du fabricant. Ensuite, une étape d'amplification par PCR à la Phire® est réalisée, avec une amorce (SPLNK1) spécifique de l'adaptateur et une amorce spécifique de l'élément *mos1*. A partir des produits de cette PCR, une deuxième amplification PCR à la Phire® est réalisée avec un autre couple d'amorce spécifique à de l'adaptateur (SPLNK2) et de l'élément *mos1*. Les produits de PCR sont déposés sur gel d'agarose, et les bandes obtenues sont découpées individuellement clonées et vérifiées par séquençage. Les clonages ont été réalisés avec le kit « Thermo Scientific CloneJet PCR Cloning », selon les recommandations du fabricant.

c) Détection des insertions de *mos1* par hybridation *in situ* sur chromosomes polytènes

Nous avons utilisé un fragment (900 paires de bases) de *mos1* (1286 paires de bases) incorporé dans un site *HindIII* du plasmide pBR 322. Les glandes salivaires de larves de stade 3 ont été écrasées dans de l'acide acétique 45%, déshydratées dans de l'éthanol 96% et séchées à l'air. Les lames ont été incubées 30 minutes dans une solution 2 x SSC à 60°C. Elles ont ensuite été incubées 5 minutes dans des bains à concentrations croissantes d'alcool (50%, 70%, 90%), puis séchées. La dénaturation des chromosomes a été effectuée par un traitement au NaOH 0,07M pendant 2 minutes. Une solution contenant la sonde ADN biotinylée-dénaturée (4 X SCC, dextran et formamide 50%) a été ajoutée aux préparations. L'hybridation a été effectuée sur la nuit à 37°C en chambre humide. Les lames ont ensuite été successivement lavées par du 2 X SCC pré-chauffé à 37°C, puis du PBS à température ambiante. Une fois la sonde hybridée sur les chromosomes, le complexe biotine-streptavidine est marqué à la peroxydase qui permet la révélation de la sonde par oxydation de diaminobenzidine en présence de peroxyde d'hydrogène (coloration brune visible en microscopie optique).

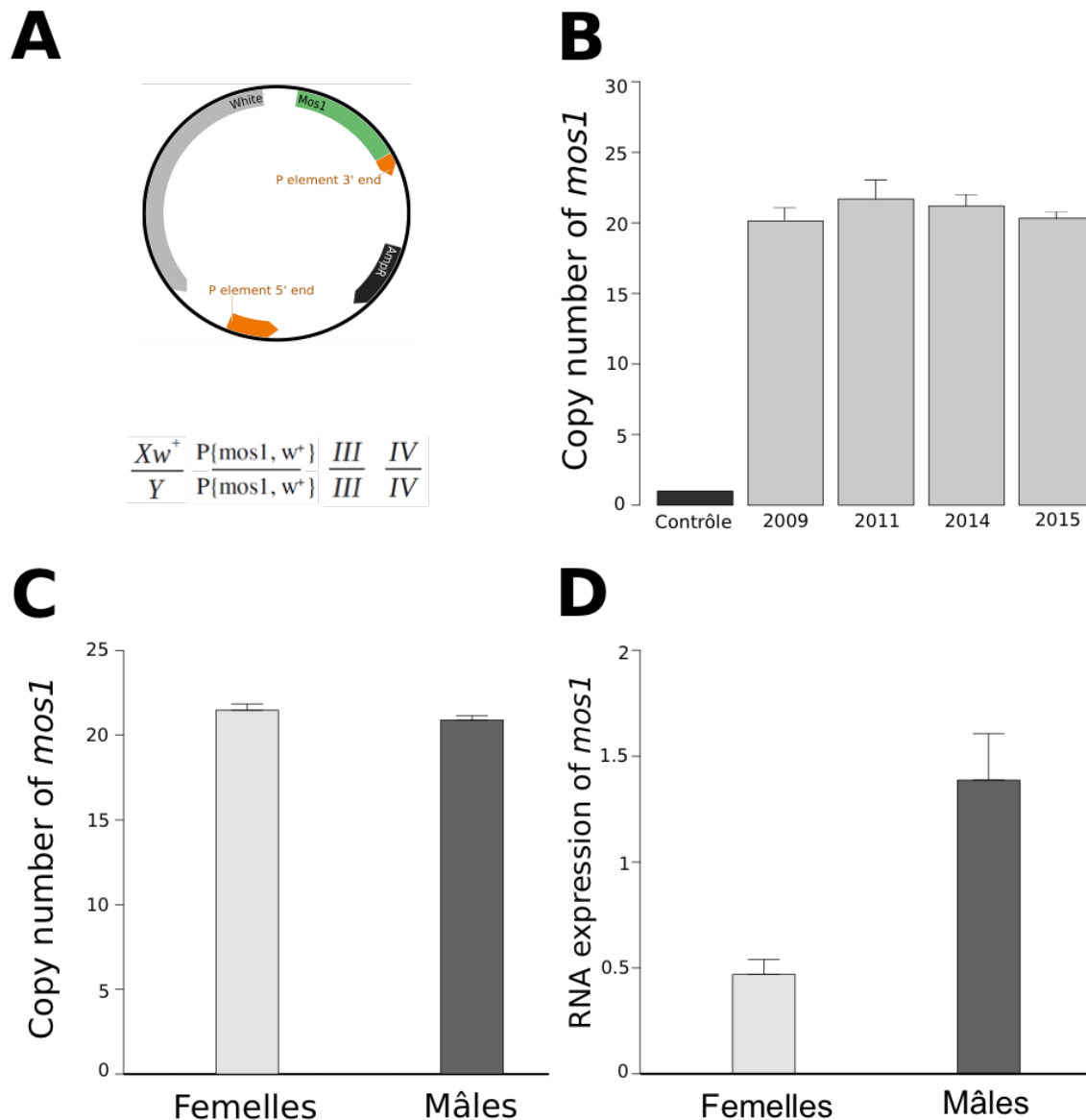


Figure 1. Caractéristiques de *mos1* depuis son introduction dans la lignée M19

La figure (A) montre le vecteur utilisé pour l'introduction de *mos1* dans le génome de la lignée M19. Le vecteur est composé de la séquence complète de *mos1* (vert), du gène *mini-white* (gris), des pieds de l'élément *P* (orange) et d'un gène de résistance à l'ampicilline (noir). La figure (B) montre la quantification par RT qPCR de *mos1* (barre grise) à plusieurs années d'intervalles (2009, 2011, 2014 et 2015). En 2004, le nombre de copies a été estimé à 20 copies par southern blot (données non présentées). Le contrôle (barre noire) est une lignée ne possédant qu'une seule copie de *mos1* immobilisée. La figure (C) montre le nombre de copies estimées par RT-qPCR chez des adultes mâles (noir) et femelles (gris). La figure (D) montre le niveau de transcription de *mos1* chez les mâles et femelles adultes. L'expression de *mos1* a été estimée par RT-qPCR, et normalisée par trois gènes de références (RPL17, Light et RP49).

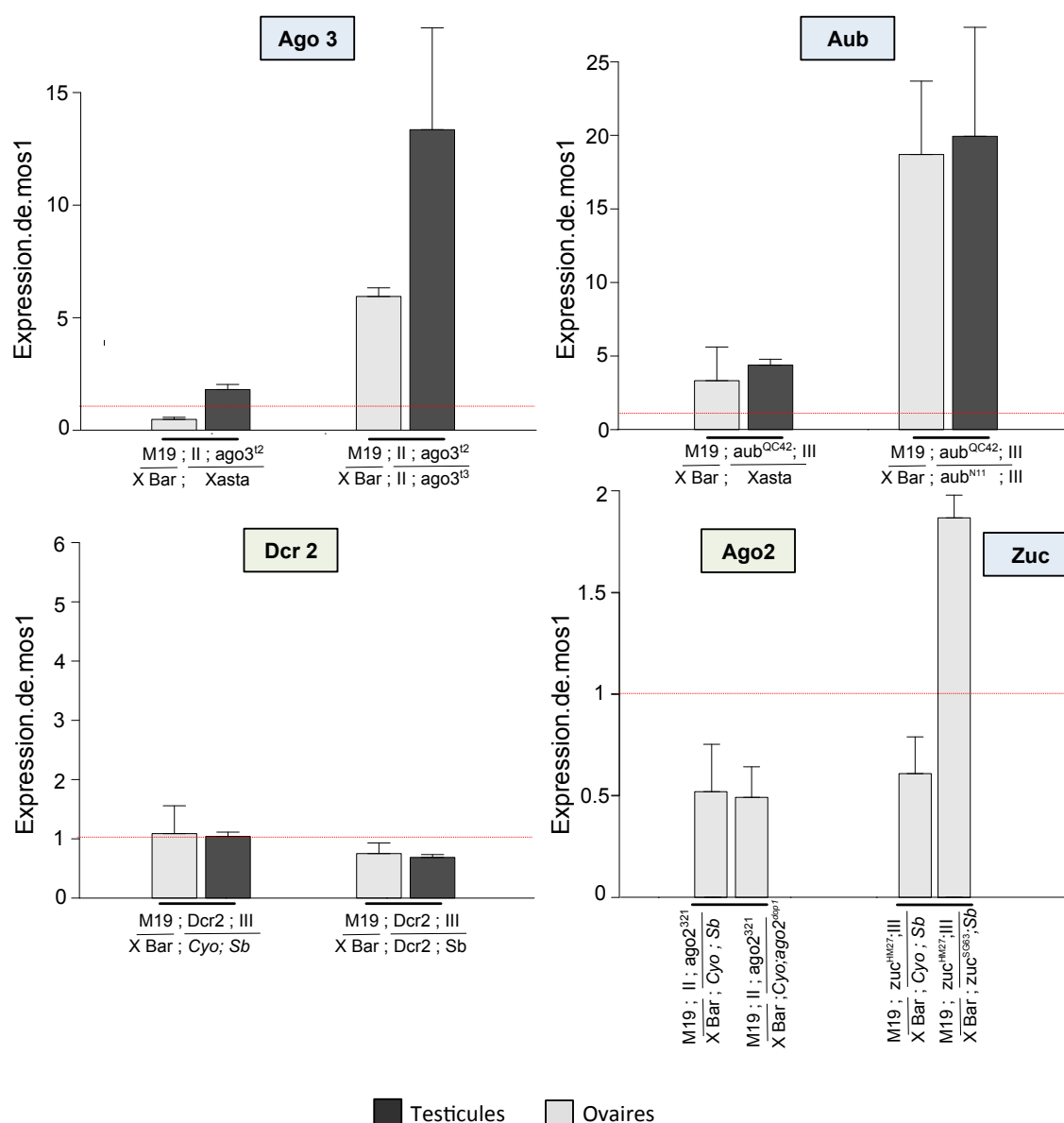


Figure 2. Étude de l'activité transcriptionnelle de *mos1* en contexte mutant

La figure 2 montre les variations d'activité transcriptionnelle de *mos1* entre mutants hétérozygotes et mutants hétéroalléliques pour des gènes associés aux voies de régulation des *piARN* et des *siARN*. La ligne rouge correspond au niveau d'expression de *mos1* dans les ovaires de la lignée M19 (souche d'origine). Les données présentées correspondent à l'écart relatif à cette mesure. Les données d'expression ont été normalisées par les gènes de ménage *his4* et *Rps29* et par la méthode des $\Delta\Delta Ct$ (cf méthodes). En gris sont présentées les données relatives à l'expression dans l'ovaire et en noir les données relatives à l'expression dans les testicules. Les gènes mutés associés à la régulation par les *piARN* sont *argonaute3* (*ago3* - A), *aubergine* (*aub* - B), et *zucchini* (*zuc* - D). Les gènes mutés associés à la régulation par les *siARN* sont *dicer2* (*dcr2* - C) et *argonaute2* (*ago2* - D). Pour chaque condition, les génotypes sont indiqués sur l'axe des abscisses.

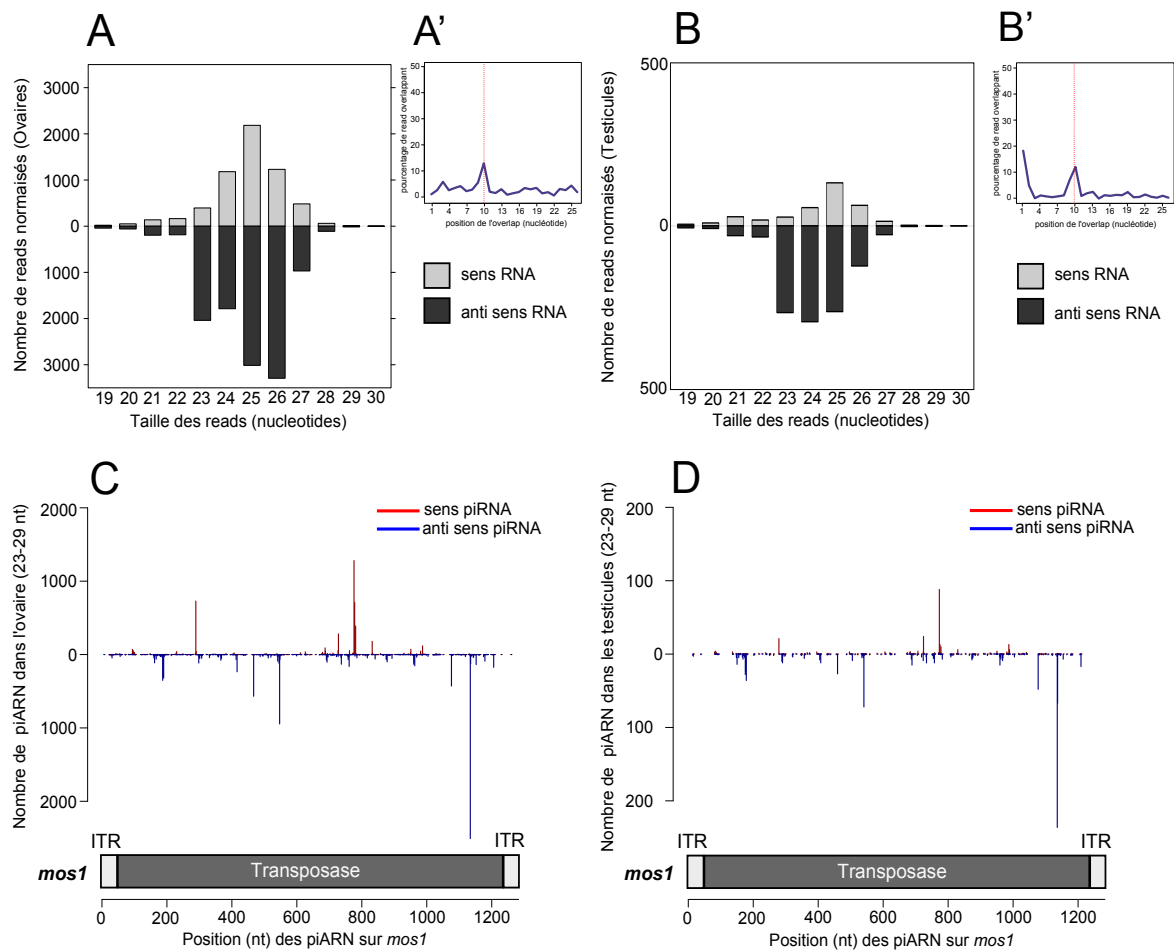


Figure 3. La production *de novo* de *piARN* dérivés de *mos1*

La figure (A) montre le nombre de petits ARN ovariens dérivés de *mos1* (axe des ordonnées) en fonction de leur taille exprimée en nucléotides (axe des abscisses). Les barres grises correspondent aux petits ARN sens et les barres noires aux petits ARN antisens. La figure (A') montre le pourcentage de *piARN* qui se chevauchent (ordonnées) en fonction de la position initiale du chevauchement (abscisses). La ligne verticale rouge montre le pic de la distribution (au nucléotide 10). La figure (B) montre le nombre de petits ARN dérivés de *mos1* produits dans les testicules (axe des ordonnées) en fonction de leur taille exprimée en nucléotides (axe des abscisses). Les barres grises correspondent aux petits ARN sens et les barres noires aux petits ARN antisens. La figure (B') montre le pourcentage de *piARN* qui se chevauchent (ordonnées) en fonction de la position initiale du chevauchement (abscisses). La ligne verticale rouge montre le pic de la distribution (au nucléotide 10). Les figures (C) et (D) montrent le nombre de *piARN* (23-29 nt) qui s'alignent sur *mos1* (ordonnées), en fonction de leur position sur la séquence l'élément (abscisses) respectivement pour les *piARN* ovariens (C) et ceux produits dans les testicules (D). L'élément *mos1* est schématisé par un rectangle avec en gris la séquence codant pour la transposase et en blanc les séquences terminales inversées répétées (ITR).

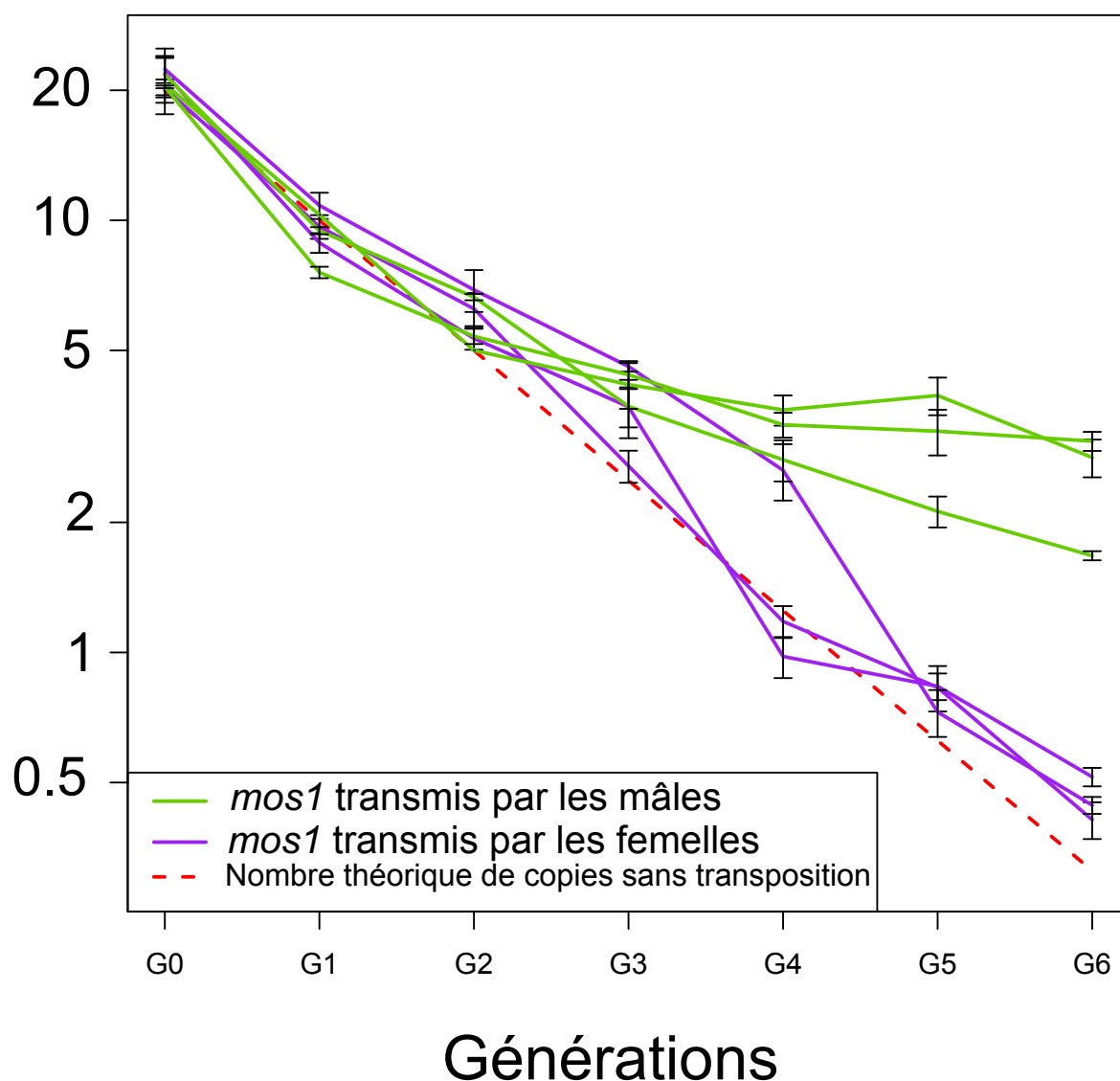


Figure 4. Rétrocroisements successifs entre la lignée M19 et des souches vides de *mos1*

La figure 4 représente le nombre de copies estimé par RT-qPCR, et présenté *via* une échelle logarithmique (axe des ordonnées), en fonction des générations successives de rétrocroisements (axe des abscisses). La ligne rouge pointillée correspond à la diminution théorique de nombre de copies de *mos1*, sous l'hypothèse que *mos1* ne transpose pas. Les courbes violettes décrivent l'évolution du nombre de copies de *mos1* au cours de 3 croisements indépendants où l'élément est systématiquement transmis par la mère. Les courbes vertes correspondent au cas où, *mos1* est systématiquement transmis par le père.

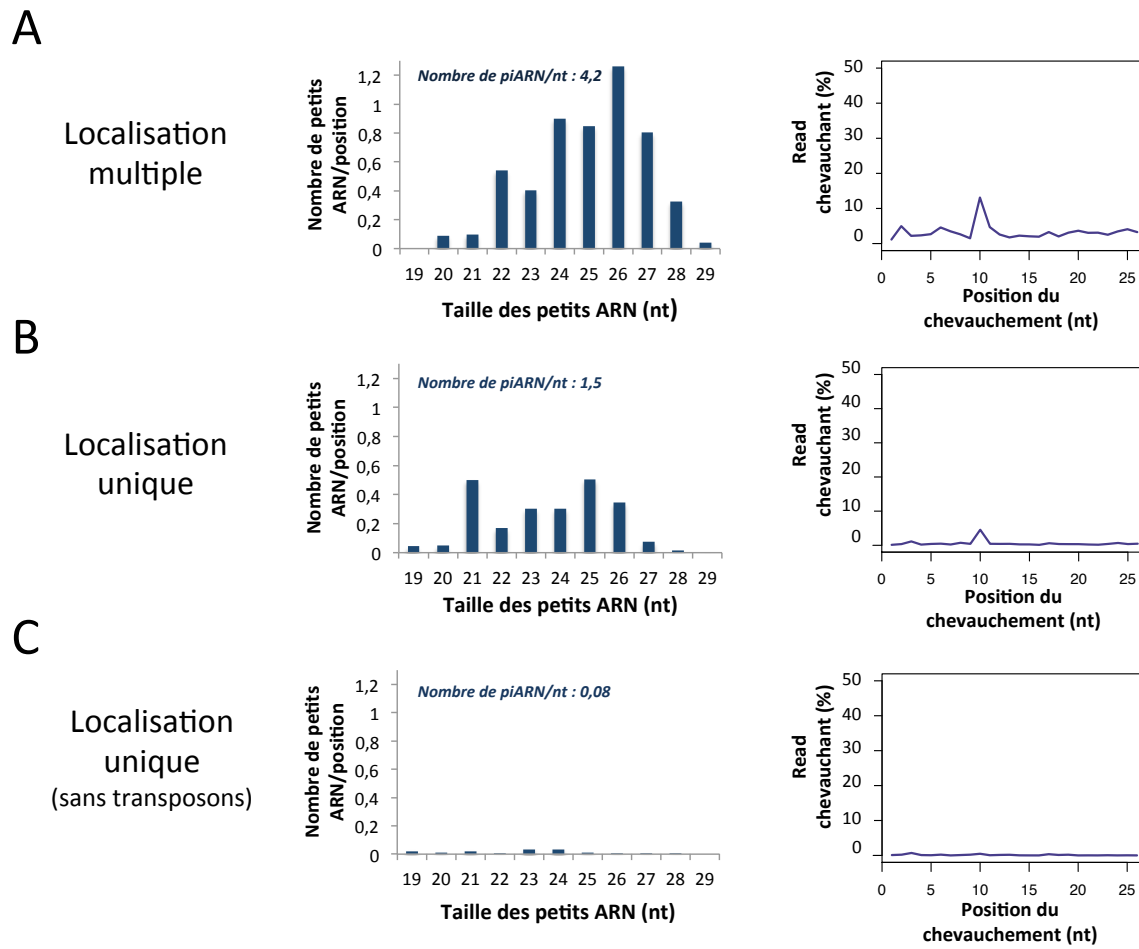


Figure 5. La production de *piARN* autour des insertions de *mos1* dans la lignée M19

Les panneaux de gauche montrent la quantité moyenne de petits ARN par nucléotide dans les séquences présentes autour des insertions de *mos1* (axes des ordonnées). Les petits ARN sont distribués selon leur taille en nucléotides (axe des abscisses). Les panneaux de droite (nombre d'appariement en fonction de la position initiale du chevauchement) représentent le pourcentage de *piARN* présentant une signature ping-pong. La figure (A) montre l'enrichissement local en petits ARN dans les régions répétées du génome de *D. melanogaster*. La figure (B) montre l'enrichissement en petits ARN dans les régions non répétées du génome. La figure (C) montre l'enrichissement en petits ARN dans les régions non répétées du génome et non apparentées à des ET.

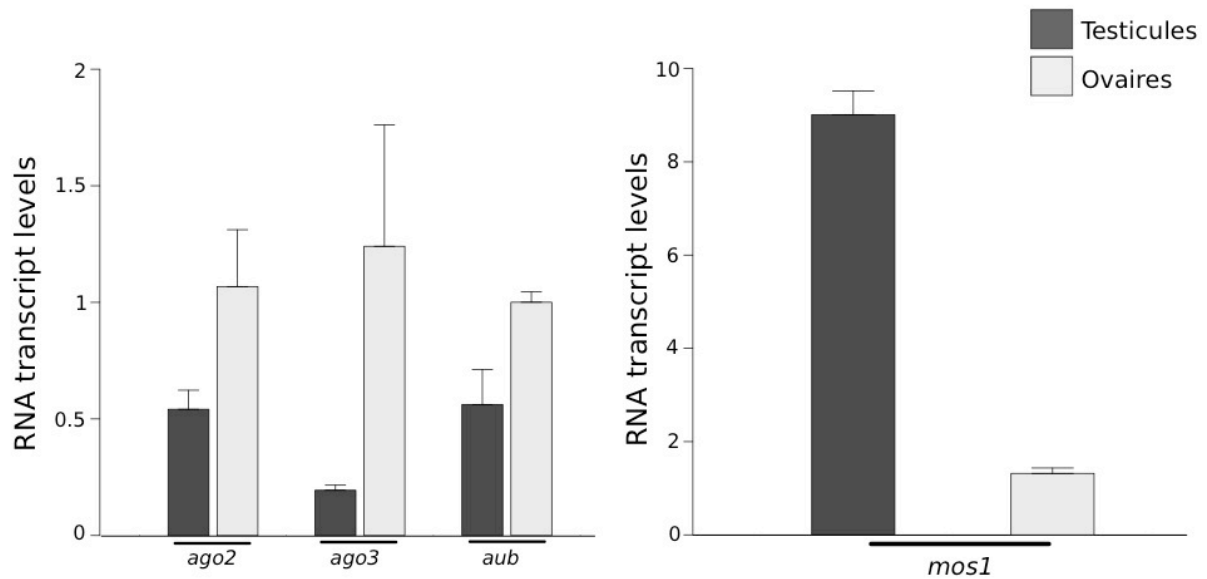


Figure Supplémentaire 1

Les niveaux d'expression des gènes *ago2*, *ago3*, *aub* (A) et de l'élément *mos1* (B) sont présentés. Les gènes de ménage utilisés pour la normalisation des données sont His4 et Rps29. La figure montre l'expression relative des gènes associés à la régulation par ARN interférents dans les ovaires (gris) et dans les testicules (noir).

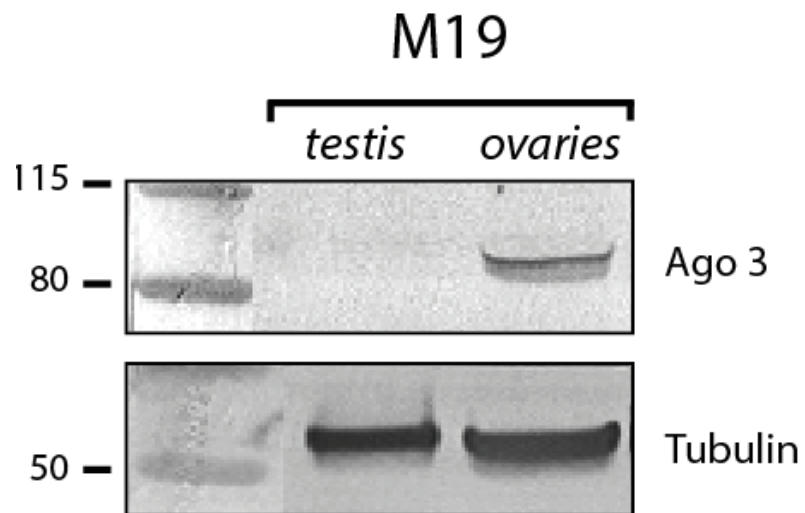


Figure Supplémentaire 2

La figure montre le niveau d'expression qualitatif de la protéine Ago3 par Western blot dans les ovaires et les testicules de la lignée M19, normalisée par la tubuline.

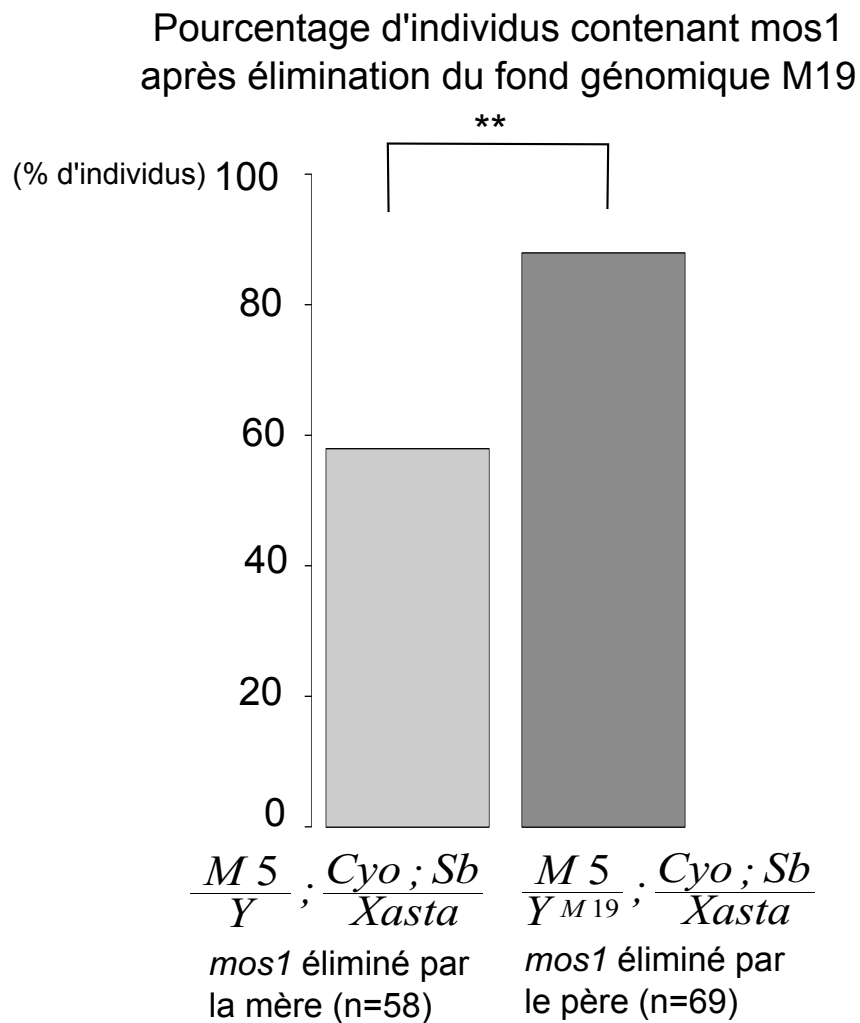


Figure Supplémentaire 3

La figure montre le pourcentage d'individus présentant au moins une copie de *mos1* (quantifiée par PCR sur gel), après remplacement total du fond génomique de la lignée M19. A gauche, lorsque le fond génomique est éliminé par la mère, à droite, lorsque le fond génomique est éliminé par le père.

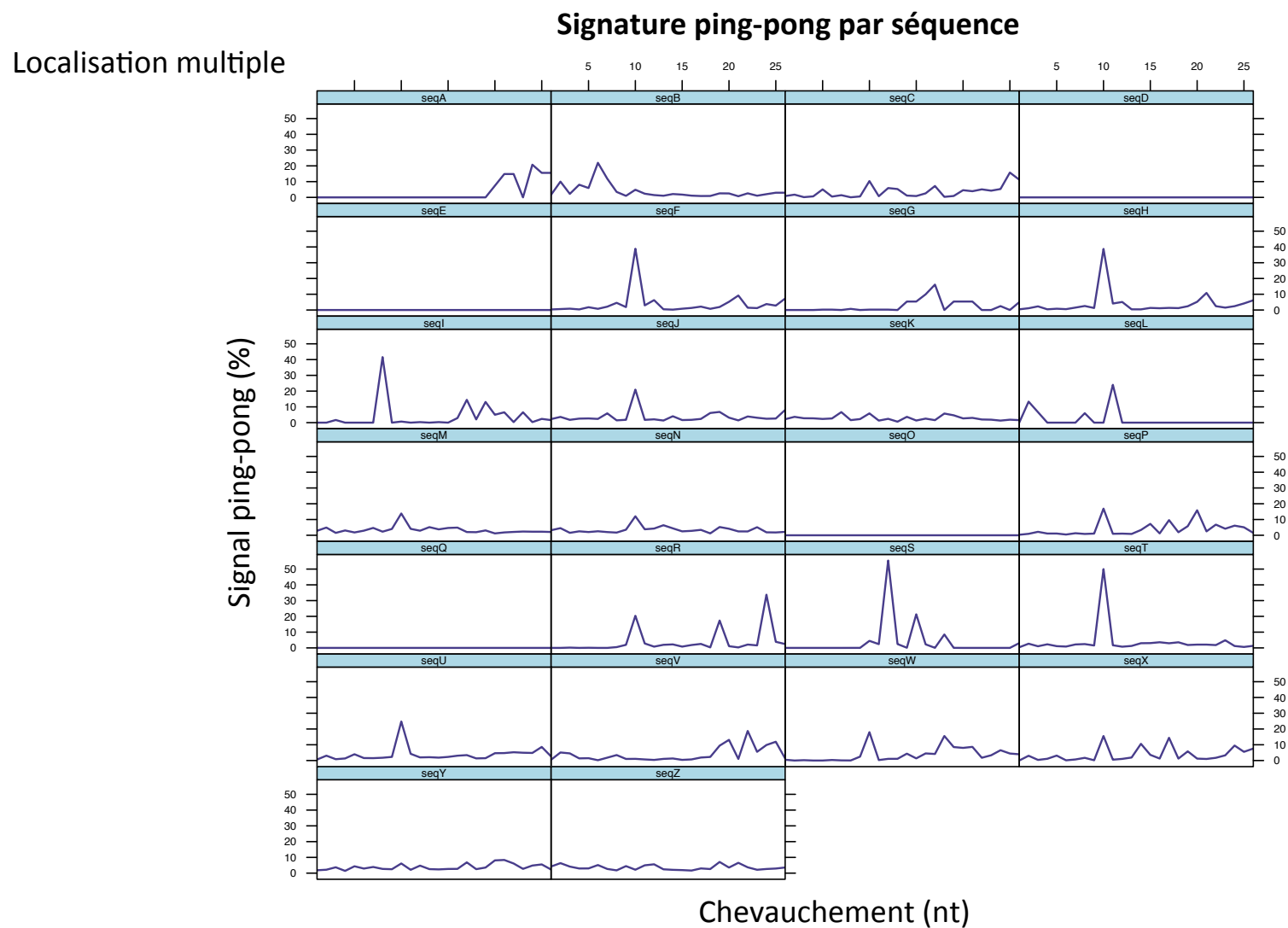
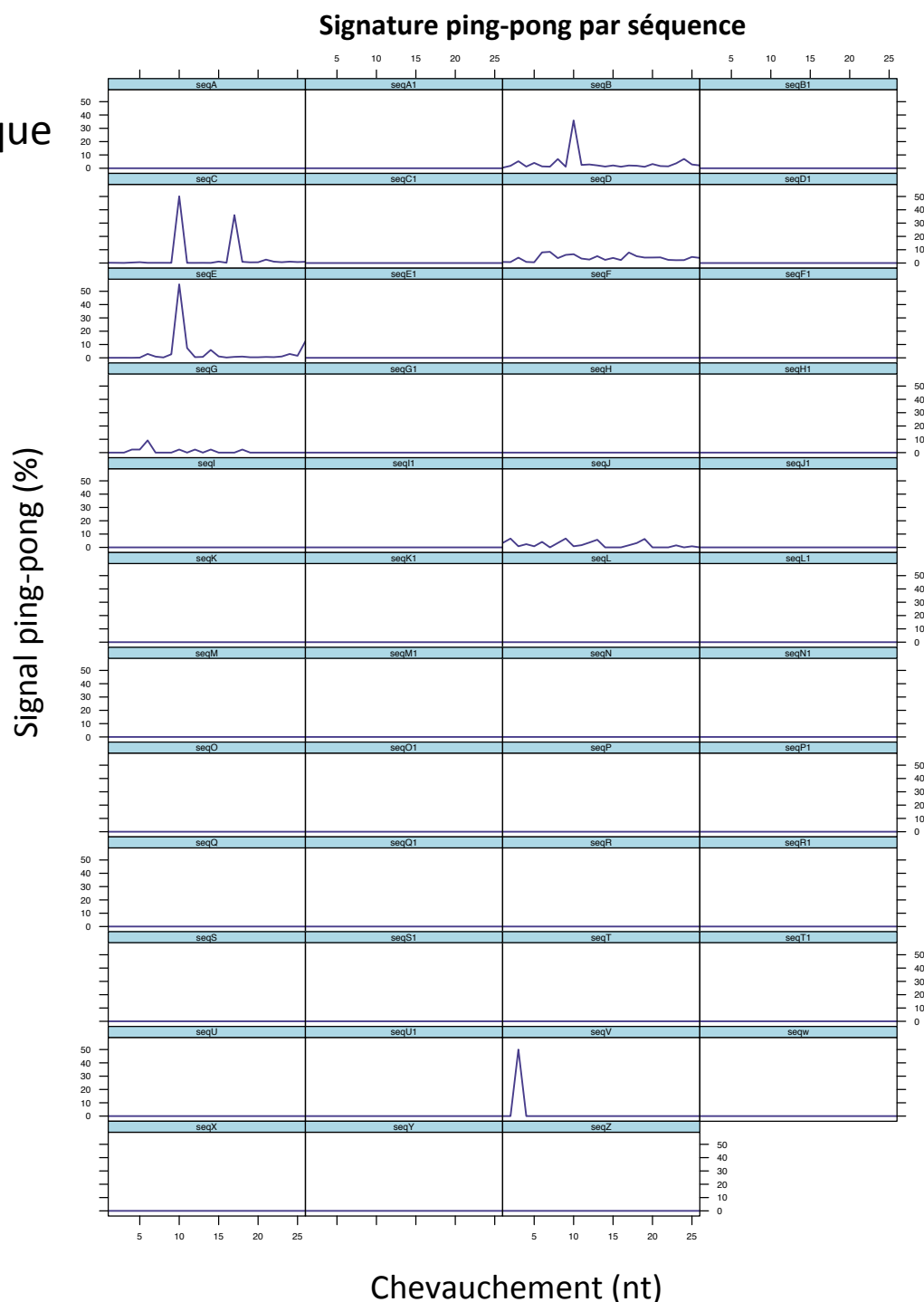


Figure Supplémentaire 4 (1/2)

Localisation unique

**Figure Supplémentaire 4 (2/2)**

La figure montre les signatures ping-pong associées aux *piARN* ovariens mappant sur chacune des régions du génome qui correspondent aux séquences flanquantes des insertions de *mos1*. L'axe des ordonnées correspond au pourcentage de reads chevauchant exprimé en fonction de la position initiale du chevauchement sur le read (axe des abscisses). Le nom des séquences (rectangle bleu), fait référence à ceux présentés dans la première colonne du tableau 2 et 3.

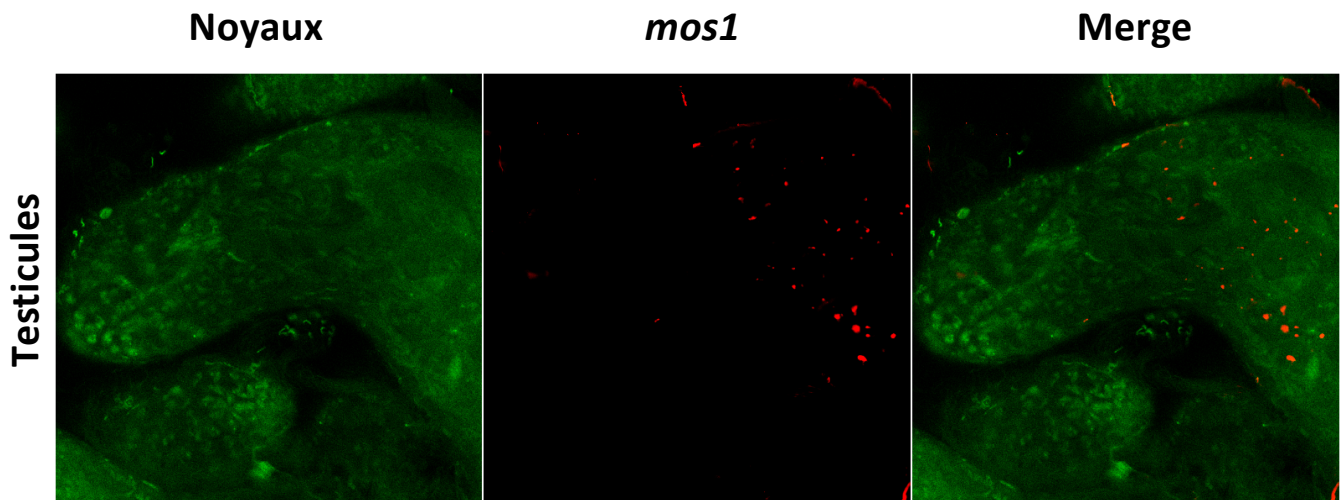


Figure Supplémentaire 5

Hybridation in situ de *mos1* dans les testicules de la génée M19. Les noyaux sont en vert et les transcrits sens de *mos1* sont en rouge.

		Chr X	Bras 2L	Bras 2R	Bras 3L	Bras 3R
Localisation sur les chromosomes	Multiples chez plusieurs individus		30F	45C	62F	90F
		-		48D	63B	85E
					77D	
	Mutiples chez le même individu	6F8A	22D	44D	75C	99C
		8D	29D	50C		92F
		9E	29F	51D		92A
		9A		59C		90D
		10F		59A		87B
				54D		87EF
				48A		
				43E		
	Unique dans un individu	15C	24D		64D	87EF
			33DE		76	84EF
			35D			



Tableau 1

Le tableau présente les insertions de *mos1* détectées par hybridation *in situ* sur chromosomes polytènes. Les colonnes correspondent à la localisation des bandes cytologiques sur chaque bras chromosomiques (X, 2L, 2R, 3L, 3R). Aucune insertion n'a été localisée sur le 4 et sur le Y. Deux individus ont été analysés, les insertions sont classées suivant leur présence dans plusieurs noyaux au sein des deux individus (partie haute du tableau), leur présence chez un seul individu mais dans plusieurs noyaux (partie intermédiaire) et celles retrouvées uniquement une fois (partie basse du tableau). La photo est un exemple de localisation des insertions de *mos1* détectées (coloration brune).

Tableau 2

Le tableau représente les séquences adjacentes des insertions de *mos1* (obtenues par PCR séquençage) que l'on peut associer à une région unique du génome. Les séquences C, G, H et I sont retrouvées à deux (C et G) ou trois (H et I) localisations génomiques distinctes.

ID	Élément transposable	Localisation génomique	Localisation chromosomique	Gène	Répétition
seqA	N/A	2L : 1886323..1886297	22B4	CG7337	unique
seqB	G2+unique	3R : 102405..101904	N/A	N/A	unique
seqC	RC1A	3L : 24935566..24935701	N/A	N/A	dupliqué
seqD	LTR_297	2R : 631532..631315	N/A	CG45781	unique
seqE	protop_B+LTR_297	2R : 631692..631317	N/A	CG45781	unique
seqF	N/A	3L 18047625..18047720	75B5	CG8127	unique
seqG	Gypsy_5	2R : 678999..678889	N/A	CG45781	dupliqué
seqH	N/A	N/A	N/A	N/A	multiple
seqI	N/A	N/A	N/A	N/A	multiple
seqJ	Gypsy_5	3L : 24911615..24912035	N/A	N/A	unique
seqK	NotoAg1	2L : 16184044..16183879	35E6	CG4894	unique
seqL	N/A	2L : 16183982..16183879	35E6	CG4894	unique
seqM	N/A	3L : 8896111..8896360	66E3	CG5751	unique
seqN	N/A	3L : 8895921..8896189+ 8896320..8896360	66E3	CG5751	unique
seqO	N/A	3L : 8896102..8896402	66E3	CG5751	unique
seqP	N/A	3L : 8895921..8896381 + 8896337..8896402	66E3	CG5751	unique
seqQ	N/A	3L : 15305526..15305261	71C2	CG7255	unique
seqR	N/A	3L : 15305825..15305271	71C2	CG7255	unique
seqS	Gypsy12-LTR	Unmapped Scaffold_8_D1580_D1567: 71895..71702	N/A	CG12423	unique
seqG1	N/A	3L:18045957..18046264	75B5	CG5751	unique
seqH1	N/A	X : 8038525..8038654	73D	N/A	unique
seqT	N/A	2R : 9083254..9082905	45A1	CG11778	unique
seqU	Harbinger-2	2R : 10701095..10701528	47B5	CG11895	unique
seqV	N/A	X : 8038654..8038519	7D3	CG2252	unique
seqW	gypsy44	3R : 30746986..30747622	100A3	CR43238	unique
seqX	N/A	2R : 7720797..7721346	43E13	CG18812	unique
seqY	Copia-43	X : 2541316..2541178	3A1	CG33950	unique
seqZ	Copia-43	X : 2541261..2541178	3A1	CG33950	unique
seqA1	Copia-43	X : 2541316..2541178	3A1	CG33950	unique
seqB1	N/A	X : 2541251..2541178	3A1	CG33950	unique
seqC1	N/A	3L : 5708466..5708327	64E2	CG34418	unique
seqD1	N/A	2L : 21573708..21573670	39E2	CG3305	unique
seqE1	N/A	X : 19360920..19361028	18C3	CG12199	unique
seqF1	N/A	2R : 7720963..7721361	43E13	CG18812	unique
seqI1	N/A	2L : 7020445..7020396	27D4	CG31628	unique
seqJ1	N/A	2L : 8489065..8489180	29D1	CG13088	unique
seqK1	N/A	2L : 8489065..8489180	29D1	CG13088	unique
seqL1	N/A	2L : 8489066..8489575	29D1	CG13088	unique
seqM1	N/A	2L : 8489066..8489180	29D1	CG13088	unique
seqN1	N/A	2L : 12689746..12689947	33F3	N/A	unique
seqO1	N/A	3L : 2781892..2782426	62F4	CG2101	unique
seqP1	N/A	3L : 3153852..3154106	63B7	CG43365	unique
seqQ1	N/A	3L : 8528807..8528460	66D5-6	CG6817	unique
seqR1	N/A	3L : 16745793..16745510	73C2	N/A	unique
seqS1	N/A	3L : 20502959..20502934	77C6	CG32425	unique
seqT1	N/A	3L : 22176487..22176367	79C3	CG12673	unique
seqU1	N/A	3R : 31266532..31266556	100B9	CG15555	unique

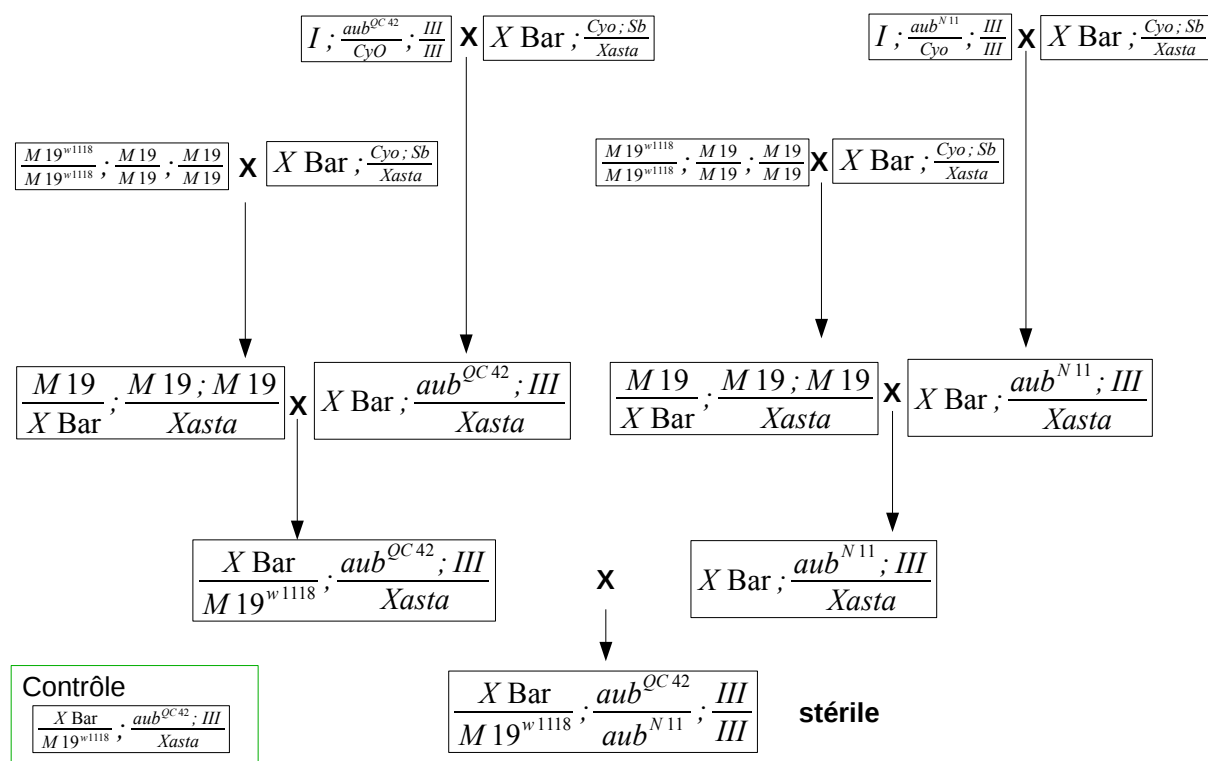
ID	ET	Localisation génomique
seqA	IVK	multiple
seqB	RT1A	multiple
seqC	RT1A	multiple
seqD	Copia-25	2L1886538 ..1886299(22B4)
seqE	Protop_B	multiple
seqF	G2	multiple
seqG	G2	multiple
seqH	G2	multiple
seqI	G2	multiple
seqJ	G2 et Doc	multiple
seqK	Doc	multiple
seqL	Gypsy-6	multiple
seqM	HMS_BEAGLE	multiple
seqN	HMS_BEAGLE	multiple
seqO	PROTOP_B	multiple
seqP	RC1A	multiple
seqQ	RC1A	multiple
seqR	PROTOP_B	multiple
seqS	FW	multiple
seqT	RC1A	multiple
seqU	RC1A	multiple
seqV	PROTOP	multiple
seqW	PROTOP_PROTOP_B	multiple
seqX	protop_B	multiple
seqY	297 et 1.76	multiple
seqZ	HMS_BEAGLE	multiple

Tableau 3

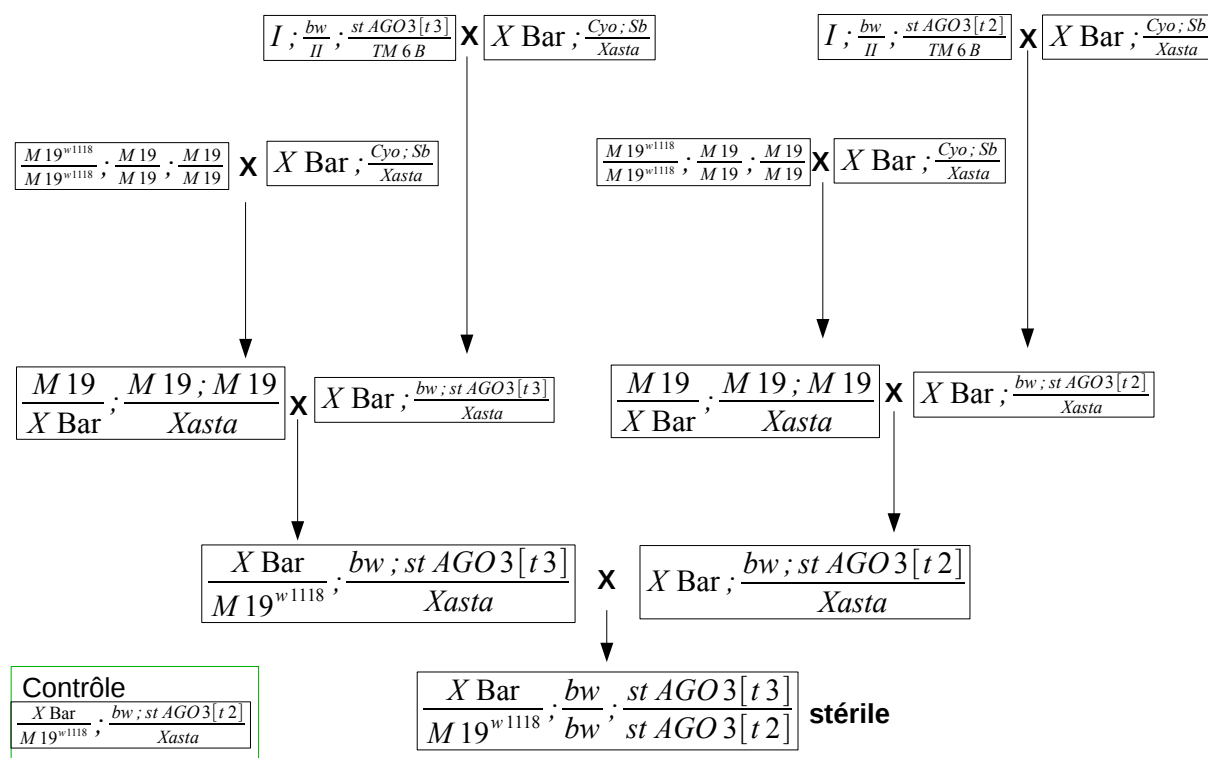
Le tableau montre les séquences adjacentes des insertions de *mosI* (obtenues par PCR séquençage) que l'on peut associer à des régions répétées du génome (uniquement des ET).

TABLEAUX DE CROISEMENTS

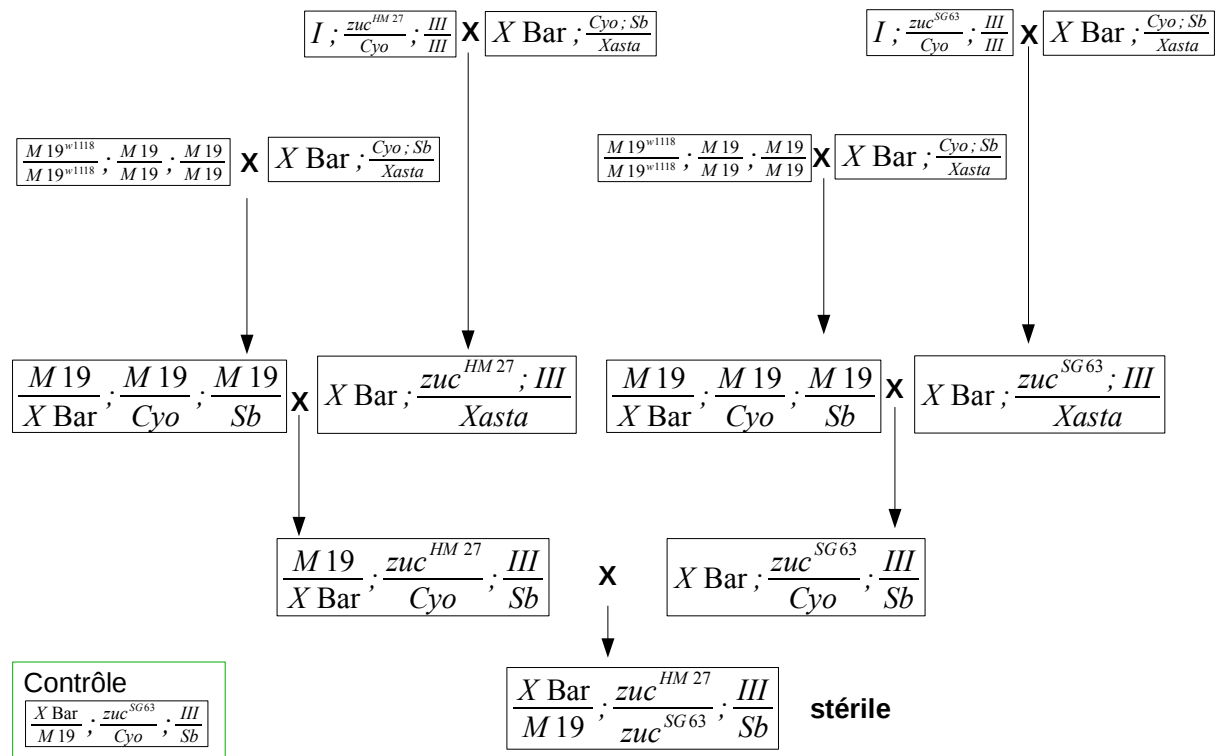
1. Croisement hétéroallélique entre la lignée M19 et les lignées Aubergine déficientes (*aub*[QC42] et *aub*[N11])



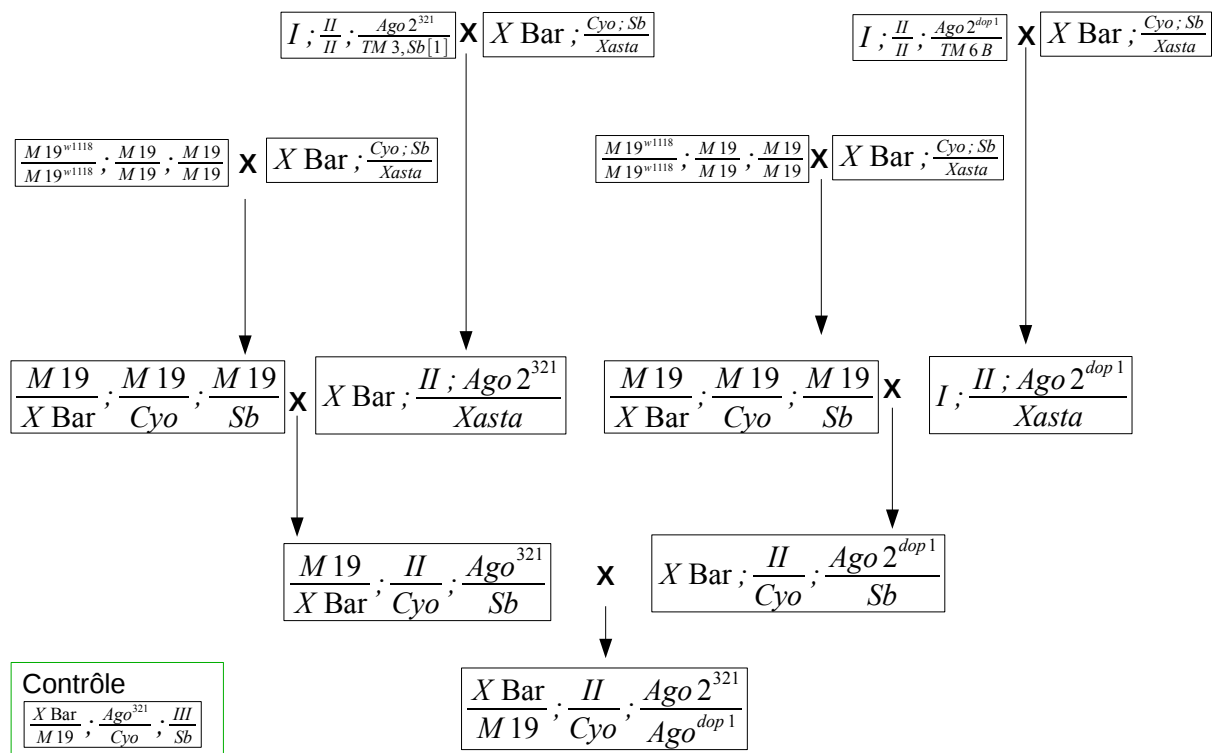
2. Croisement hétéroallélique entre la lignée M19 et les lignées Ago3 déficientes (*Ago3*[t2] et *Ago3*[t3])



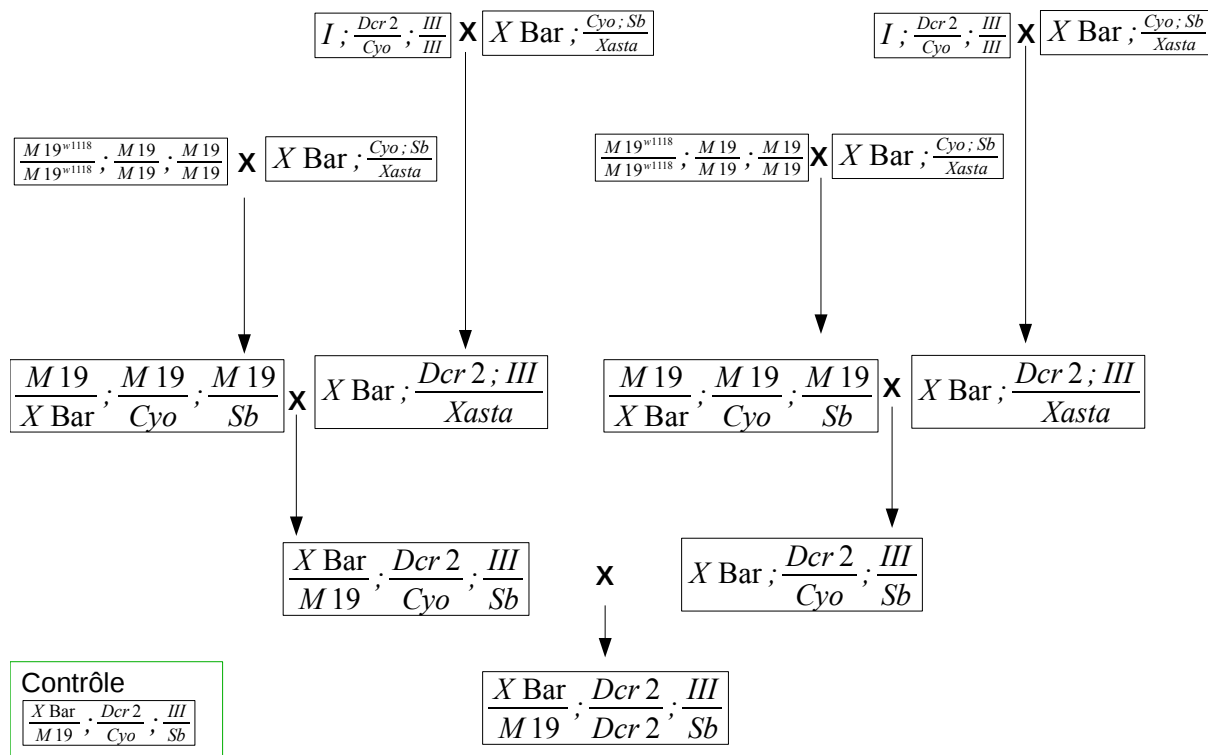
3. Croisement hétéroallélique entre la lignée M19 et les lignées Zucchini déficientes (*zuc* [HM27] et *zuc* [SG63])



4. Croisement hétéroallélique entre la lignée M19 et les lignées Ago2 déficientes (*Ago2* [321] et *Ago2* [dop1])

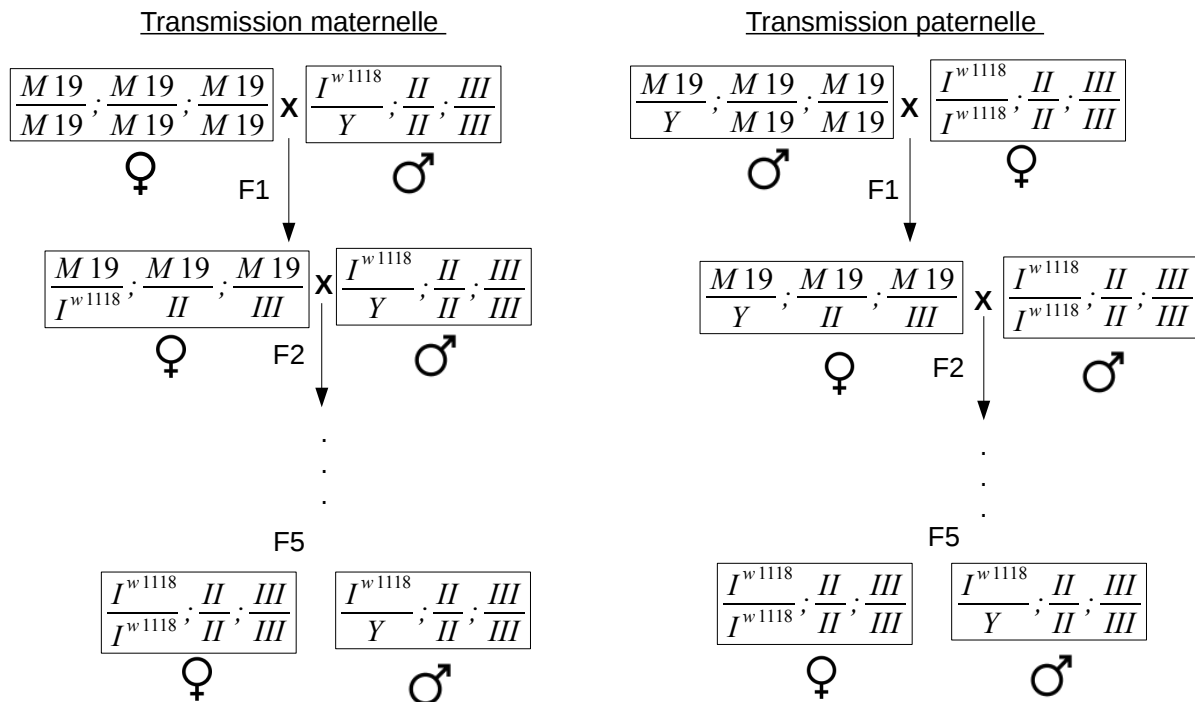


5. Croisement hétéroallélique entre la lignée M19 et les lignées Dicer2 déficientes (dcr2 [L811fs])



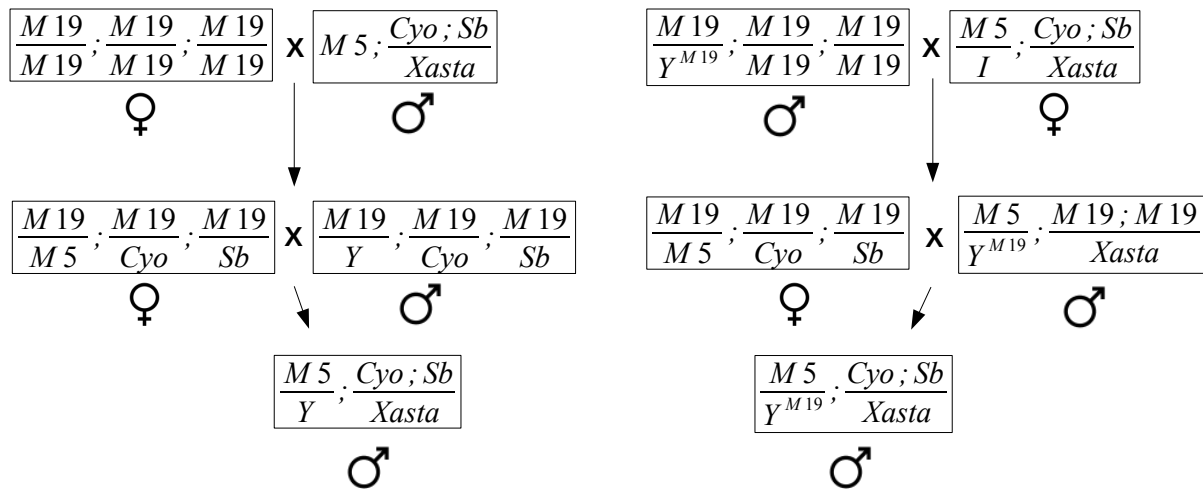
6. Rétrocroisements successifs entre la lignée M19 et la lignée w1118 (sans l'élément mos1)

L'élément *mos1* est transmis soit par les mères soit par les pères de la lignée M19



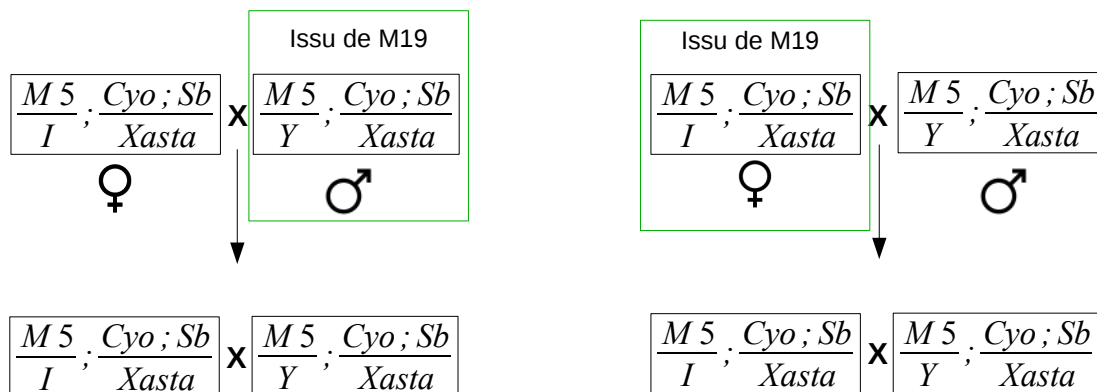
Quantification du nombre de copies à chaque génération

7. Croisement entre la lignée M19 et les lignées $M5; \frac{Cyo; Sb}{Xasta}$ sans l'élément *mos1*



Présence ou absence de *mos1* chez chaque individu de la F2

8. Elimination du Y M19 (contrôle du croisement $M19 \times M5; \frac{Cyo; Sb}{Xasta}$)



Quantification du nombre de copies

Bibliographie

- Andersen, C.L., J.L. Jensen, and T.F. Orntoft. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64:5245-5250.
- Aravin, A., and T. Tuschl. 2005. Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett.* 579:5830-5840.
- Aravin, A.A., and G.J. Hannon. 2008. Small RNA silencing pathways in germ and stem cells. *Cold Spring Harb Symp Quant Biol.* 73:283-290.
- Bagijn, M.P., L.D. Goldstein, A. Sapetschnig, E.M. Weick, S. Bouasker, N.J. Lehrbach, M.J. Simard, and E.A. Miska. 2012. Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science.* 337:574-578.
- Blumenstiel, J.P., and D.L. Hartl. 2005. Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. *Proc Natl Acad Sci U S A.* 102:15965-15970.
- Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell.* 128:1089-1103.
- Brennecke, J., C.D. Malone, A.A. Aravin, R. Sachidanandam, A. Stark, and G.J. Hannon. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science.* 322:1387-1392.
- Brunet, F., F. Godin, C. Bazin, and P. Capy. 1999. Phylogenetic analysis of Mos1-like transposable elements in the Drosophilidae. *J Mol Evol.* 49:760-768.
- Bucheton, A., R. Paro, H.M. Sang, A. Pelisson, and D.J. Finnegan. 1984. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. *Cell.* 38:153-163.
- Capy, P. 1998. Evolutionary biology. A plastic genome. *Nature.* 396:522-523.
- Chung, W.J., K. Okamura, R. Martin, and E.C. Lai. 2008. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol.* 18:795-802.
- Czech, B., C.D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J.A. Wohlschlegel, R. Sachidanandam, G.J. Hannon, and J. Brennecke. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature.* 453:798-802.
- de Setta, N., M.A. Van Sluys, P. Capy, and C.M. Carareto. 2009. Multiple invasions of Gypsy and Micropia retroelements in genus *Zaprionus* and *melanogaster* subgroup of the genus *Drosophila*. *BMC Evol Biol.* 9:279.
- de Vanssay, A., A.L. Bouge, A. Boivin, C. Hermant, L. Teyssset, V. Delmarre, C. Antoniewski, and S. Ronsseray. 2012. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature.* 490:112-115.
- Farazi, T.A., S.A. Juranek, and T. Tuschl. 2008. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development.* 135:1201-1214.
- Giraud, T., and P. Capy. 1996. Somatic activity of the mariner transposable element in natural populations of *Drosophila simulans*. *Proc Biol Sci.* 263:1481-1486.
- Grentzinger, T., C. Armenise, C. Brun, B. Mugat, V. Serrano, A. Pelisson, and S. Chambeyron. 2012. piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res.* 22:1877-1888.
- Gunawardane, L.S., K. Saito, K.M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M.C. Siomi. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science.* 315:1587-1590.
- Hain, D., B.R. Bettencourt, K. Okamura, T. Csorba, W. Meyer, Z. Jin, J. Biggerstaff, H. Siomi, G. Hutvagner, E.C. Lai, M. Welte, and H.A. Muller. 2010. Natural variation of the amino-terminal glutamine-rich domain in *Drosophila argonaute2* is not associated with developmental defects. *PLoS One.* 5:e15264.
- Hamilton, A.J., and D.C. Baulcombe. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science.* 286:950-952.
- Handler, D., D. Olivieri, M. Novatchkova, F.S. Gruber, K. Meixner, K. Mechtler, A. Stark, R. Sachidanandam, and J. Brennecke. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* 30:3977-3993.
- Ito, H., H. Gaubert, E. Bucher, M. Mirouze, I. Vaillant, and J. Paszkowski. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature.* 472:115-119.

- Ivancevic, A.M., A.M. Walsh, R.D. Kortschak, and D.L. Adelson. 2013. Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. *Bioessays*. 35:1071-1082.
- Josse, T., L. Teyssset, A.L. Todeschini, C.M. Sidor, D. Anxolabehere, and S. Ronsseray. 2007. Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet*. 3:1633-1643.
- Juliano, C., J. Wang, and H. Lin. 2011. Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annu Rev Genet*. 45:447-469.
- Kawamura, Y., K. Saito, T. Kin, Y. Ono, K. Asai, T. Sunohara, T.N. Okada, M.C. Siomi, and H. Siomi. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature*. 453:793-797.
- Kelleher, E.S., and D.A. Barbash. 2013. Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. *Mol Biol Evol*. 30:1816-1829.
- Kelleher, E.S., N.B. Edelman, and D.A. Barbash. 2012. Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol*. 10:e1001428.
- Ketting, R.F., T.H. Haverkamp, H.G. van Luenen, and R.H. Plasterk. 1999. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell*. 99:133-141.
- Khurana, J.S., J. Wang, J. Xu, B.S. Koppetsch, T.C. Thomson, A. Nowosielska, C. Li, P.D. Zamore, Z. Weng, and W.E. Theurkauf. 2011. Adaptation to P element transposon invasion in Drosophila melanogaster. *Cell*. 147:1551-1563.
- Kidwell, M.G., and J.B. Novy. 1979. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: Sterility Resulting from Gonadal Dysgenesis in the P-M System. *Genetics*. 92:1127-1140.
- Klenov, M.S., O.A. Sokolova, E.Y. Yakushev, A.D. Stolyarenko, E.A. Mikhaleva, S.A. Lavrov, and V.A. Gvozdev. 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci U S A*. 108:18760-18765.
- Koundakjian, E.J., D.M. Cowan, R.W. Hardy, and A.H. Becker. 2004. The Zuker collection: a resource for the analysis of autosomal gene function in Drosophila melanogaster. *Genetics*. 167:203-206.
- Lee, Y.C. 2015. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in Drosophila melanogaster. *PLoS Genet*. 11:e1005269.
- Lee, Y.S., and R.W. Carthew. 2003. Making a better RNAi vector for Drosophila: use of intron spacers. *Methods*. 30:322-329.
- Lee, Y.S., K. Nakahara, J.W. Pham, K. Kim, Z. He, E.J. Sontheimer, and R.W. Carthew. 2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*. 117:69-81.
- Loreto, E.L., C.M. Carareto, and P. Capy. 2008. Revisiting horizontal transfer of transposable elements in Drosophila. *Heredity (Edinb)*. 100:545-554.
- Malone, C.D., and G.J. Hannon. 2009. Small RNAs as guardians of the genome. *Cell*. 136:656-668.
- Mejlumian, L., A. Pelisson, A. Bucheton, and C. Terzian. 2002. Comparative and functional studies of Drosophila species invasion by the gypsy endogenous retrovirus. *Genetics*. 160:201-209.
- Meyer, W.J., S. Schreiber, Y. Guo, T. Volkmann, M.A. Welte, and H.A. Muller. 2006. Overlapping functions of argonaute proteins in patterning and morphogenesis of Drosophila embryos. *PLoS Genet*. 2:e134.
- Morris, C.A., E. Benson, and H. White-Cooper. 2009. Determination of gene expression patterns using in situ hybridization to Drosophila testes. *Nat Protoc*. 4:1807-1819.
- Murchison, E.P., P. Stein, Z. Xuan, H. Pan, M.Q. Zhang, R.M. Schultz, and G.J. Hannon. 2007. Critical roles for Dicer in the female germline. *Genes Dev*. 21:682-693.
- Nagao, A., T. Mituyama, H. Huang, D. Chen, M.C. Siomi, and H. Siomi. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. *RNA*. 16:2503-2515.
- Olovnikov, I., S. Ryazansky, S. Shpiz, S. Lavrov, Y. Abramov, C. Vauray, S. Jensen, and A. Kalmykova. 2013. De novo piRNA cluster formation in the Drosophila germ line triggered by transgenes containing a transcribed transposon fragment. *Nucleic Acids Res*. 41:5757-5768.
- Pfaffl, M.W., A. Tichopad, C. Prgomet, and T.P. Neuvians. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnol Lett*. 26:509-515.

- Picot, S., G.L. Wallau, E.L. Loreto, F.O. Heredia, A. Hua-Van, and P. Capy. 2008. The mariner transposable element in natural populations of *Drosophila simulans*. *Heredity (Edinb)*. 101:53-59.
- Rand, T.A., S. Petersen, F. Du, and X. Wang. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*. 123:621-629.
- Rozhkov, N.V., A.A. Aravin, R. Sachidanandam, G.J. Hannon, O.N. Sokolova, E.S. Zelentsova, N.G. Shostak, and M.B. Evgen'ev. 2010a. The RNA interference system differently responds to the same mobile element in distant *Drosophila* species. *Dokl Biochem Biophys*. 431:79-81.
- Rozhkov, N.V., A.A. Aravin, E.S. Zelentsova, N.G. Schostak, R. Sachidanandam, W.R. McCombie, G.J. Hannon, and M.B. Evgen'ev. 2010b. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA*. 16:1634-1645.
- Rozhkov, N.V., N.G. Schostak, E.S. Zelentsova, I.A. Yushenova, O.G. Zatsepina, and M.B. Evgen'ev. 2013. Evolution and dynamics of small RNA response to a retroelement invasion in *Drosophila*. *Mol Biol Evol*. 30:397-408.
- Rozhkov, N.V., E.S. Zelentsova, N.G. Shostak, and M.B. Evgen'ev. 2011. Expression of *Drosophila virilis* retroelements and role of small RNAs in their intrastrain transposition. *PLoS One*. 6:e21883.
- Rubin, G.M., and A.C. Spradling. 1982. Genetic transformation of *Drosophila* with transposable element vectors. *Science*. 218:348-353.
- Schaack, S., C. Gilbert, and C. Feschotte. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 25:537-546.
- Schupbach, T., and E. Wieschaus. 1991. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics*. 129:1119-1136.
- Shpiz, S., S. Ryazansky, I. Olovnikov, Y. Abramov, and A. Kalmykova. 2014. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the *drosophila* germline. *PLoS Genet*. 10:e1004138.
- Sijen, T., and R.H. Plasterk. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature*. 426:310-314.
- Siomi, M.C., K. Sato, D. Pezic, and A.A. Aravin. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol*. 12:246-258.
- Svoboda, P. 2004. Long dsRNA and silent genes strike back: RNAi in mouse oocytes and early embryos. *Cytogenet Genome Res*. 105:422-434.
- Tabara, H., M. Sarkissian, W.G. Kelly, J. Fleenor, A. Grishok, L. Timmons, A. Fire, and C.C. Mello. 1999. The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell*. 99:123-132.
- Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 3:RESEARCH0034.
- Watanabe, T., A. Takeda, T. Tsukiyama, K. Mise, T. Okuno, H. Sasaki, N. Minami, and H. Imai. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev*. 20:1732-1743.
- Watanabe, T., Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M.A. Surani, Y. Sakaki, and H. Sasaki. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*. 453:539-543.
- Wilson, J.E., J.E. Connell, and P.M. Macdonald. 1996. aubergine enhances oskar translation in the *Drosophila* ovary. *Development*. 122:1631-1639.
- Woodruff, R.C., J.N. Thompson, Jr., J.S. Barker, and H. Huai. 1999. Transposable DNA elements and life history traits: II. Transposition of P DNA elements in somatic cells reduces fitness, mating activity, and locomotion of *Drosophila melanogaster*. *Genetica*. 107:261-269.
- Yang, N., and H.H. Kazazian, Jr. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*. 13:763-771.
- Zilberman, D., X. Cao, and S.E. Jacobsen. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*. 299:716-719.

Article n°2

Transcriptional polymorphism of *pi*RNA regulatory genes underlies the *mariner* activity in *D. simulans* testes

Transcriptional polymorphism of *pi*RNA regulatory genes underlies the *mariner* activity in *D. simulans* testes.

BASTIEN SAINT-LEANDRE, ISABELLE CLAVEREAU, AURELIE HUA-VAN and PIERRE CAPY

Laboratoire Evolution, Génomes, Comportement, Ecologie CNRS, Univ. Paris-Sud, IRD, Université Paris-Saclay, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette cedex, France.

Abstract

Transposable Elements (TE) are selfish DNA segments that contribute to the structural and functional evolution of genomes. They are also important sources of genetic variability between populations and species. During colonization of new areas, natural populations have to deal with changing and stressful environments, and TEs can be useful “tools” in the adaptation process. In this general context, the activity (copy number, transcriptional and excision rate) of the *mariner mos1* element was estimated in 19 natural populations of *D. simulans*. It is shown (1) that *mos1* expression is always higher and more variable in testes than in ovaries; (2) that *mos1* activity is higher in colonizing populations compared to the sub-Saharan African ones (ancestral populations); (3) that *mos1* variations in transcript levels and copy number are negatively correlated to transcriptional variations of *pi*RNA genes, *Aubergine* and *Argonaute3*. Furthermore, *mos1* levels of expression in testes highly contrast with the low expression patterns of *Ago3*. These results strongly suggest that the expression polymorphism of *pi*RNA genes could be responsible for the *mos1* variations, first between male and female germlines and second, according to the status of natural populations (invasive or not). These results provide new perspectives about TE and *pi*RNA genes' co-evolution in the *Drosophila* germlines.

Keywords: transposable element activity, natural population, colonization, *mariner*, *Drosophila simulans*.

Correspondence: Pierre Capy, Fax: 33 (1) 69.82.37.36; E-mail: pierre.capy@egce.cnrs-gif.fr.

Introduction

Phenotypic divergence (between species) and polymorphism (between populations) are consequences of variations at different biological levels (genetic, epigenetic, development) and of effects of changing environmental conditions. In this respect, it has been admitted for a long time that transposable elements (TEs) play an important role both in terms of structural and functional evolution of the genomes (McClintock 1984). For example, it is shown that most of the genome size variation in eukaryotes is related to TE abundance (Kalendar *et al.* 2000; Kidwell 2002; Kooke *et al.* 2015; Lee & Kim 2014; Wessler 2006). In addition, mobile elements have been repeatedly captured by their host for new cellular functions, as illustrated by domestication events (Sinzelle *et al.* 2009).

McClintock (1984) proposed that mobilization of TEs could have been selected as a response of the genome to stressful environmental conditions. Indeed, several sets of evidence suggest a relationship between stress and the TE mobilization (Arnault & Dufournel 1994; Capy *et al.* 2000; Hagan *et al.* 2003; Kalendar *et al.* 2000; Kidwell & Lisch 1997; Vieira *et al.* 1999). In particular, the awakening of TE activity has been observed in natural populations that colonized different ecological niches. In *Drosophila* species, this was suggested both for *D. melanogaster* and *D. simulans* (Biemont *et al.* 2003; Vieira & Biemont 2004). In Hawaiian *Drosophila*, the copy number of *Uhu* and *LOA* elements is higher in populations that colonized the youngest islands compared to populations of older islands (Wisotzkey *et al.* 1997). Similarly, Labrador *et al.* (Labrador *et al.* 1998) have shown that the *Oswaldo* element in

D. buzzatii exhibits a higher copy number in populations of the colonized area (Iberian Peninsula) compared to that of the ancestral area (Argentina). Nevertheless, it remains difficult to determine whether TE reactivation in colonizing populations is due to an environmental selection or a simple founder effect (Garcia Guerreiro *et al.* 2008).

D. melanogaster and *D. simulans* are cosmopolitan sibling species with similar geographic distribution and demographic history, such as an expansion and a worldwide colonization from the ancestral African area, but with a different colonizing history (Capy & Gibert 2004; David & Capy 1988; Irvin *et al.* 1998; Lachaise *et al.* 1988; Lachaise & Silvain 2004; Nolte & Schlotterer 2008). For *D. melanogaster*, the main hypothesis related to its colonization is an adaptation to human modified environments 10,000 to 15,000 years ago with the emergence of agriculture, followed by the colonization of Europe and Eurasia, and later of the American and Australian continents (David & Capy 1988).

In contrast, the worldwide colonization of *D. simulans* is thought to be more recent (a few centuries: see Capy & Gibert 2004 for review). The ancestral geographic area of *D. simulans* is thought to be East Africa, including the Islands of the Indian Ocean close to Africa (Madagascar, Reunion, Mayotte), from where it has spread to new habitats (Dean & Ballard 2004; Kopp *et al.* 2006; Lachaise & Silvain 2004; Schoefer & Schlotterer 2006). Several works showing the very low differentiation between North African and European populations compared to those observed between tropical African and European ones support the species expansion out of Africa through the “Nile route” (Bouletreau-Merle *et al.* 2003; Chakir *et al.* 2002; Picot *et al.* 2008; Schoefer & Schlotterer 2006). *D. simulans* stands out from *D. melanogaster* by its recent worldwide expansion from European populations (Begun & Aquadro 1995; Kopp *et al.* 2006; Singh *et al.* 1987; Veuille *et al.* 2004) to North America, Asia and Oceania. Moreover, *D. simulans* is still an active invading species that recently colonized the Japanese islands in the 1970s (Watanabe & Kawanishi 1976) and Australia (Bock & Parsons 1981).

During the colonization process, populations of *D. melanogaster* and *D. simulans* got used to their new environments. In this respect, clinal variations (latitudinal and altitudinal) have been described for many quantitative traits probably due to climatic selection (Arthur *et al.* 2008; Capy *et al.* 1994; David & Bocquet 1975; Hoffmann & Weeks 2007; Vieira & Biemont 2004; Weeks *et al.* 2002), but also for the distribution of TEs (Boussy *et al.* 1998; Picot *et al.* 2008; Vieira & Biemont 2004).

This could be due to the fact that TEs can play a role of mutator agents in adaptation to local environments, as shown in *Drosophila* (Gonzalez *et al.* 2008; Gonzalez & Petrov 2009), especially in adaptation to temperate environments (Gonzalez *et al.* 2010).

Insertions or recruitment of TEs as *cis*-regulatory elements at a key adaptive locus has been selected many times during evolution, and is still likely underestimated due to technical limitations for the identification of *de novo* TE insertions (Casacuberta & Gonzalez 2013). Nevertheless, few examples of adaptive insertions have been described, such as the induction of insecticide resistance (DDT) in both *D. melanogaster* and *D. simulans* (Daborn *et al.* 2002; Schlenke & Begun 2004). Furthermore, TEs are generally assumed to be slightly deleterious and thus removed or maintained at low frequencies in populations (Kofler *et al.* 2015).

Independently of TE insertions maintained or lost by natural selection or by genetic drift, the host regulatory machinery is of central importance to limit TE family expansion. In the germline of metazoans, the *pi*RNA pathway mediates both transcriptional and post-transcriptional silencing of TEs. The regulatory machinery drives highly specific inhibition by the production of antisense *pi*RNA that target TEs sense mRNA (Pelisson *et al.* 2007; Saito & Siomi 2010; Senti & Brennecke 2010; Siomi *et al.* 2011; Vagin *et al.* 2009). In *Drosophila* ovaries, *Aub* and *Ago3* are the core genes of the *pi*RNA amplification process, and are essential to the germline silencing of the most active TE families (Brennecke *et al.* 2007; Gunawardane *et al.* 2007). These two genes and many of the *pi*RNA pathway effector proteins exhibit a

strong signature of recurrent adaptive substitutions (Kolaczowski *et al.* 2011; Obbard *et al.* 2009a; Obbard *et al.* 2009b). It is thus assumed that evolution of these genes is rapid and driven by an arms race between TE activity and host regulatory machinery. However, only coding sequences were analyzed in these studies and transcriptional variations of these genes have been poorly investigated so far. Supporting the idea that *piRNA* genes are under diversifying selection, only one recent work has shown a high level of transcriptional variation among *D. simulans* strains (Fablet *et al.* 2014).

In the present work, we will focus on the germinal regulation of the *mos1* DNA transposon initially described in *D. mauritiana* (Medhora *et al.* 1988). This element belongs to the *mauritiana* subfamily of *mariner* elements (Rouault *et al.* 2009) and is present in several species of the *melanogaster* complex. The somatic activity of the element can be estimated through a phenotypic excision assay that relies on the mobilization of an inactive copy by the active copies present in the populations. This test revealed that *mos1* somatic activity is highly variable between natural populations of *D. simulans*. These variations in activity seemed to be organized in a latitudinal cline from Africa to Europe (Giraud & Capy 1996) or in Australia (Russell & Woodruff 1999), but this hypothesis remained dubious due to the opposite tendencies observed. A more recent work (Picot *et al.* 2008) concluded from a detailed re-analysis of these data that the variations observed were probably due to the population dynamics and more precisely to the route followed during their colonization history. Indeed, the somatic activity was globally positively correlated with the invasive distance to the geographical origin area of the species (East Africa).

The *mos1* element is therefore a good model to investigate the dynamics of a transposable element during the invasion process. Moreover, the tools available in *D. simulans* allow us to dissect the molecular mechanisms involved in the activity regulation of this element. Here, we analyzed the variation of *mos1* activity in female and male germlines of *D. simulans* in connection with the expression of *piRNA* genes. We observed that in all populations and species, the *mos1* activity is higher in testes than in ovaries. Surprisingly, we observed the opposite pattern for *Ago3*. Interestingly, in *D. simulans*, the activity of the element is higher in colonizing populations compared to the non-invasive ones. Moreover, we found the expression level of *mos1* activity to be negatively correlated to the expression of *Aub* and *Ago3* in ovaries.

Materials and methods

Fly stocks

D. simulans populations were collected from geographically distinct areas. More precisely they are from Indian Ocean islands (Mayotte, Reunion, Seychelles), Africa (Nairobi, Kenya; Harare [Zimbabwe]; Yaoundé [Cameroon]; Sao Tome; Mohammedia and Marrakech [Morocco]; Alexandria [Egypt], Europe (Bordeaux and Draveil, France); Asia (Tokyo, Kyoto, Fukuoka and Higa, Japan) and America (Igassu [Brazil]; Zarate [Argentina]; Saint-Martin Island [Caribbean]). These populations were maintained in the laboratory as isofemales lines or mass cultures with a minimum of 50 pairs from one generation to the next. Flies were bred at 25°C on axenic medium.

All other species used in the present work belong to the *melanogaster* subgroup. All of them are endemic to Africa or islands in the Indian Ocean close to Africa. These species are: *D. yakuba* from Ivory Coast, *D. teissieri* from Mount Selinda (Zimbabwe); *D. mauritiana* from Mauritius Island and *D. sechellia*, from Mahé Island (Seychelles archipelago).

All populations and species were collected between 2000 and 2013. For a subset of

them, population/sample characteristics (only for the percentage of mosaic males [PMM] and the transcriptional activity) were tested at different moments to verify the stability. As already shown (by Picot *et al.* 2008), the PMM is stable when the samples are kept several years under laboratory conditions or when they are trapped at different periods. We also validated the stability of transcriptional activity in populations from Mayotte, Reunion and Sao Tomé collected at different times between 2000 and 2013 (data not shown).

RNA extraction, synthesis of cDNA and quantitative RT-PCR analysis

30 to 50 individuals were used for each experimental condition. Tissues were dissected in PBS, conserved in RA1 buffer (Macherey-Nagel) and stored at -80°C until use. Samples were grinded with steel beads (3mm) before total RNA isolation performed with the NucleoSpin® RNA kit (Macherey-Nagel) following the manufacturer's protocol. RNA concentrations were estimated by measuring their absorbance at 260nm on a NanoDrop 2000 spectrophotometer. Quality control of the RNA was performed with the Experion™ Automated Electrophoresis System (Bio-Rad). Only samples with higher quality index (RQI > 7) were kept. To avoid contamination by genomic DNA, a second DNA digestion was performed following the Turbo DNA-free™ kit recommendations (Ambion). We used the iScript™ cDNA synthesis kit (Bio-Rad) to synthesize the cDNA from RNA (RNA concentration = 1000ng/μl). The cDNA samples were stored at -20°C until used.

Gene expressions were quantified using the CFX96™ Real-Time PCR Detection System (Bio-Rad) and with iQ SYBR® Green supermix (Bio-Rad). These expressions were normalized with reference genes, with the $\Delta\Delta$ Cq method. Primers for RT-qPCR were designed on Primer3 software (<http://bioinfo.ut.ee/primer3/>) from the annotated gene sequences in the *D. simulans* reference genome (www.flybase.org). To avoid quantification biases, stable housekeeping genes were pre-selected from the RefGenes database (<http://www.refgenes.org/rg/index.jsp>) and the Sex Bias Database (www.sebida.de). Primers for RT-qPCR were designed on Primer3 software (<http://bioinfo.ut.ee/primer3/>) from the annotated gene sequences in the *D. simulans* reference genome (www.flybase.org). For each experimental set-up, we normalized gene expression with the more stable housekeeping genes according to three statistical software programs: BestKeeper (Pfaffl *et al.* 2004); NormFinder (Andersen *et al.* 2004); geNorm (Vandesompele *et al.* 2002). All quantifications were based on the mean of three biological replicates.

Copy number quantification

Genomic DNA was extracted with the Chelex 100 (BioRad) according to the manufacturer's protocol. Estimation of *mos1* copy number was done using Real Time qPCR on genomic DNA. To this end, the number of *mos1* amplification cycles (CT) was normalized by the *white peach* strain that contains only one copy of the *mos1* element. Estimation of copy number is the mean of three biological replicates, and performed by $\Delta\Delta$ Cq method, using RPL17q2 as reference gene. The primers used were designed on conserved regions of all known copies of the *mos1* subfamily (Table S1).

Phenotypic excision test (PMM)

The *mos1* somatic activity was estimated from a test-cross between males of the population tested and females of the *white-peach* reference strain, as described in (Giraud & Capy 1996). The somatic activity is based on the percentage of males with mosaic eyes in F1.

Statistical Analyses

All statistical analyses were performed using the R software (Ihaka & Gentleman 1996). According to the number of tested effects we performed one-way or two-way ANOVA. The

different linear models were selected for minimum AIC (Akaike Index Criteria from the R package). The relationship between *mos1* copy number and the invasive distance from the origin area of the species was analyzed using the Mantel test (the `mantel.rtest` function of the `ade4` package, with 9999 permutations [Chessel *et al.* 2004]). The relationship between expressions of *mos1* and those of regulatory genes was also analyzed from a Principal Component Analysis (`ade4` package, `dudi.pca` function).

In situ hybridization

RNA FISH was performed to localize the transcripts of transposable elements in ovaries and testes. To this end, 50 pairs of testes were dissected into 10mM Tris HCl, 47mM NaCl, 183mM KCl, pH 6.8, and the RNA-FISH experiment was carried out (according to the procedure described by Morris *et al.* 2009). Similarly, 20 pairs of ovaries were dissected into PBS before proceeding to the RNA-FISH experiment (according to the protocol described by Akkouche *et al.* 2012). The probe was synthesized with the TA cloning Kit (Invitrogen) and the primer was designed according to the manufacturer's protocol (Fmos195T7 = ggtaatacgaactcactataggAGAGCAAACGCGGACAGTAT).

Western Blot

We quantified *Ago3* levels of proteins by western blot. Fifty pairs of testes and thirty pairs of ovaries were lysed in 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Triton-X100, 5% glycerol, pH 7.5. Protein levels were quantified by Bradford assay (Biorad Protein Assay) and an identical quantity was loaded on a gel (35µg). SDS-PAGE was performed using NuPAGE 4-12% Bis-Tris gels (Life Technologies) in MES buffer (MES 50mM, Tris-Base 50mM, SDS 0.1%, EDTA 1mM, pH 7.3). Proteins were transferred using I-blot system (Life Technologies). *Ago3* antibody was provided by Brennecke *et al.* 2007. Dilution of the *Ago3* was 1:2000. We then used AP-coupled antibodies (Promega) for protein revelation.

Results

It was previously shown that *mariner mos1* somatic excision activity estimated from a phenotypic assay and copy number (as determined by Southern blots) both vary widely across species of the *melanogaster* subgroup and between natural populations (Giraud & Capy 1996; Picot *et al.* 2008). In *D. simulans*, it was observed that African populations (sub-Saharan) usually have low somatic activities and low copy numbers compared to invasive ones. In particular, for non-African populations, a significant correlation was detected between the somatic activity level and the geographical distance (in km) from the ancestral area (East Africa). These results suggested an effect of new ecosystem colonization on the activity of the *mos1* element (Picot *et al.* 2008).

However neither of these two indices is informative regarding the involvement of genomic regulation on *mos1* activity, especially in the germline. Indeed, the phenotypic assay estimates only somatic activity in adult males, while Southern blots have usually been done on adult flies of both sexes, and are relatively imprecise. We thus developed qPCR-based new approaches in order to estimate more precisely copy number and transcriptional activity variations across species or populations, to characterize the spatio-temporal expression of *mos1* within individuals, and to check for potential relationships between copy number, transcriptional level and somatic activity.

Estimation of mos1 transcriptional activity during development and between tissues

Variations of *mos1* transcript levels were quantified by RT-qPCR. A first analysis of *mos1* RNA transcript levels was done on adults (females and males separately) of five species

of the *melanogaster* subgroup (Fig. S1). It was shown that the *mos1* transcript levels are highly variable between the five species. These variations are comparable to the variations of the copy number previously estimated by different works (Fig. S1; (Brunet *et al.* 1996; Capy *et al.* 1992; Capy *et al.* 1991; Jacobson *et al.* 1986)). More surprisingly, all tested species displayed a similar pattern characterized by a significantly high level of *mos1* expression in males compared to females ($p < 5\%$).

We then better characterized the expression profile on populations of *D. simulans*, the species showing the highest sexual dimorphism regarding *mos1* expression, and for which we had extensive data about somatic activity in natural populations. For three different populations, RNA transcripts were quantified all through the life cycle, from egg to adult, with sexing performed from the 3rd instar larval stage. In the Fukuoka population, while basal low expression was observed during the earlier stages, an elevated transcription level was detected in males only, at the old pupae stage, and maintained in adults (Fig. S2). A similar pattern was observed in the two other populations of *D. simulans*, Higa (Japan) and Mohammedia (Morocco) (Data not shown).

Variations of *mos1* expression in testes

Since the metamorphosis leads to the formation of adult structures, the existence of female/male differences were checked for in both somatic (carcasses) and germinal (gonads) tissues of 19 *D. simulans* natural populations. The results clearly show a female-male difference in gonads for all populations (Fig. 1), while no differences are observed from the comparison between female and male carcasses (data not shown). The *mos1* over-expression in testes compared to ovaries varies from 2.5 times to 110.8 times more (Fig. 1). Interestingly, only the Higa population exhibits relatively high levels of transcription in ovaries compared to other populations, which will be discussed later in the light of further observations. Gonads contain both somatic and germline cells, so we performed RNA-FISH assays on the Higa population, to determine whether *mos1* expression is restricted to one kind of tissue or not. The *mos1* expression was detectable in the nucleus of germinal nurse cells in ovaries, as well as in the germinal spermatogonial cells of testes (Fig. S3).

The *mos1* activity increase in colonizing populations

mos1 activity can also be characterized by measuring the somatic excision rate (PMM based on the phenotypic test), or the average copy number (qPCR). We analyzed the relationship between these two indices and the transcription level (in testis) for 17 *D. simulans* natural populations. Results are summarized in Fig. 2a and 2b, and clearly show that the three variables are not independent. The copy number is correlated to the excision rate ($r=0.75$; $p<0.001$) and to the *mos1* transcription levels in testes ($r=0.80$; $p<0.001$). Interestingly, the *mos1* transcription levels in ovaries do not show any relationship with *mos1* copy number and excision rate (data not shown), suggesting that solely *mos1* transcription in testes is related to the element activity in natural populations.

Fig. 2c shows that *mos1* copy number increases with colonization distance from the ancestral geographic area of the species, assumed to be in East Africa and nearby islands in the Indian Ocean (Lachaise *et al.* 1988). The relationship is statistically significant ($r=0.64$ $p<0.01$; Mantel test $p<0.02$), suggesting that *mos1* was probably mobilized during the colonization process. Indeed, *mos1* activity is significantly higher in colonizing populations than in African (Sub Saharan) populations (Fig. 3). For instance, excision rate and copy number are on average 5 times higher in non-African populations compared to Sub-Saharan ones. Moreover, several environmental factors (latitude, longitude, mean temperature) have been investigated and none of them seems to be related to *mos1* variation of activity (data not shown). Although other environmental factors cannot be excluded, these results reinforce the idea that *mos1* has been mobilized during the colonization process, and could be linked to the demographic

history populations and/or to unknown environmental constraints (Picot *et al.* 2008).

In addition, it is known that interference RNA mechanisms are key processes that transcriptionally silence TEs in *Drosophila* ovaries. Here, we observed that *mos1* transcription in ovaries is several orders of magnitude smaller than transcription in testes. These results suggest that *mos1* transcription in ovaries is subject to a regulatory mechanism that represses its transcriptional expression, but not in testes. We thus looked for the variation of several interference RNA genes in order to find out if they explain *mos1* transcriptional variations.

Relationship between *mos1* and the *piRNA* pathway

RNA silencing pathways are known to post-transcriptionally repress TEs. The impact of expression variations of RNA silencing genes on TE activity has not been addressed yet, although the expression of genes involved in *piRNA* pathway is highly variable in *D. simulans* (Fablet *et al.* 2014). In addition to the *piRNA* pathway, *siRNA* regulatory machinery was shown to be involved in germline TE silencing (Rozhkov *et al.* 2010). Here, we quantified transcriptional expression of the major RNA silencing genes of both pathways in three populations (Higa, Fukuoka and Mohammedia) that display high range of variations for the three indices of *mos1* activity. More precisely, we considered RNA transcript levels of *Argonaute 3* (*Ago3*) and *Aubergine* (*Aub*) for the *piRNA* (*piwi* RNA) pathway; *Argonaute2* (*Ago2*) and *Dicer2* (*Dcr2*) for the *siRNA* (small interfering RNA) pathway; and *mos1* characteristics (copy number and transcript level). All transcript levels have been measured in somatic and germinal tissues of both sexes.

We summarized the results in a PCA (Principal Component Analysis) combining both regulatory genes and *mos1* indices (Fig. 4a). The first two axes account for 70% of the total variance (axis 1 = 40% and axis 2 = 30%). There is no clear differentiation between germinal and somatic tissues on the two axes. However, axis 1 clearly distinguished two groups: in one hand, the Higa population and in the other hand, the Fukuoka and Mohammedia populations. Interestingly, the two groups correspond to high and low levels of *mos1* activity respectively. Projection of variables on the first two axes is given (Fig. 4b). On the first axis, an opposition is observed between the activity of the genes involved in the *piRNA* pathway (*Ago3* and *Aub*) and the *mos1* status (copy number and activity). Moreover, the genes involved in the *siRNA* pathway (*Ago2* and *Dcr2*) do not contribute to the axis and are negatively correlated to the second axis. Together Fig. 4a and 4b show that *mos1* copy number and RNA transcript levels are negatively correlated to the expression genes of the *piRNA* pathway. In other word, the higher the activity of TE, the lower the activity of *piRNA* genes. More precisely, activities of *Ago3* and *Aub* are high in ovaries of the Fukuoka and Mohammedia populations, in agreement with the weakness of *mos1* transcription levels. On the contrary, the Higa population, which exhibits a high level of *mos1* activity, presents a low level of *Ago3* and *Aub* transcripts in both germlines. We confirmed *Ago3* transcriptional variations at the protein level by a western blot (Fig. S4), showing that *Ago3* is not detectable in testes and less expressed in Higa ovaries compared to Fukuoka. These results support the hypothesis that *mos1* might be under *piRNA* regulation in the female germline. Curiously, while Higa and Fukuoka populations are two recent colonizing populations of Japan (Watanabe & Kawanishi 1976), they display highly contrasted patterns for *mos1* and *piRNA* genes transcription levels, which suggests that transcriptional variation can arise rapidly. However, the relationship between *mos1* and *piRNA* genes in ovaries remains to be tested all over the population panel.

***Ago3* transcriptional variations in gonads**

As shown by the PCA analysis, the *mos1* activity seems affected at two levels by *piRNA* gene expression, first by strong male *versus* female germline-biased expression, second by transcriptional polymorphism between populations. The RNA transcript levels of *Ago3* were thus quantified in gonads of individuals from the 19 natural populations (Fig. 5). Interestingly,

RNA transcript levels of *Ago3* in ovaries are several orders of magnitude greater than in testes. The sexual differences are always female biased and vary from 3 times (Sao Tomé Island) up to 36 times (Fukuoka) compared to the testes. Indeed, transcript levels of *Ago3* in the germlines are in complete opposition to the *mos1* expression (Fig. 1). In parallel, we observed the transcriptional activity of the *Ago3* gene to be highly polymorphic between natural populations. From an ANOVA (not shown), a significant part of *Ago3* variations is explained by the origin of the population. We thus calculated correlations between *mos1* and *Ago3* expressions in testes and ovaries of populations coming from the Sub-Saharan countries or from other parts of the world (Fig. 6). In testes, no significant correlations can be observed either in colonizing or non-colonizing populations, confirming that *mos1* is not silenced by *pi*RNA pathway in the male germline. In ovaries of Sub-Saharan populations, characterized by low *mos1* activity, we did not observe any relationship with *Ago3* variations. However, in colonizing populations that display higher levels of *mos1* activity, we observed a significant negative correlation ($p = 0.013$, $n = 12$), indicating that the *mos1* regulation in ovaries might have been achieved in response to its global activation.

Discussion

McClintock first suggested that activation of TEs could be a genome response to changing environment. Indeed, TEs are often reactivated by stress (Makarevitch *et al.* 2015; Mourier *et al.* 2014). In this respect, colonization of new ecological niches is a populational challenge. Several examples in *Drosophila* species (mainly *D. melanogaster* and *D. simulans*) from comparisons between ancestral and colonizing populations support directly or indirectly this hypothesis. For example, the copy number and/or the transcriptional activities of some TEs were found to be higher in colonizing populations, suggesting their reactivation (Labrador *et al.* 1998; Nardon *et al.* 2005; Vieira & Biemont 2004; Vieira *et al.* 1999; Wisotzkey *et al.* 1997). In *D. melanogaster* an increase of genome size in old colonizing populations was also observed. However, this effect was not observed for other cosmopolitan species such as *D. simulans*, *D. subobscura* and *D. ananassae* (Nardon *et al.* 2005).

In parallel, the colonization process often involves repeated bottlenecks that could increase drift effects, leading to a stochastic loss or fixation of some TEs insertions (Fablet *et al.* 2014; Labrador *et al.* 1998). For active elements, this phenomenon is balanced by the recurrent generation of new insertions that are found at very low frequencies in the population (Kofler *et al.* 2015).

In colonizing populations, several TE insertions were found in high frequencies. They are usually associated to a strong selective sweep, suggesting an adaptive impact on the genome (Gonzalez *et al.* 2008; Gonzalez *et al.* 2009). More precisely, it seems that the increase of TE frequency arises “*sharply after the spread out of Africa*”. The authors suggested an active role in local adaptation to temperate climates. However, the best-demonstrated examples of adaptive insertions are the TE-induced resistance to pesticides that independently occurred in both *D. melanogaster* and *D. simulans* (Aminetzach *et al.* 2005; Chung *et al.* 2007; Daborn *et al.* 2002; Schlenke & Begun 2004).

In this work, it is observed that the *mos1* activity, estimated from copy number (PCR), RNA transcript levels (RT-qPCR) and somatic excision rates (PMM), increases in colonizing populations compared to Sub-Saharan ones. This finding is in agreement with the hypothesis of TE revival due to the colonizing process. Here, we could not detect any clear relation between *mos1* activity and environmental factors, such as temperature, latitude or longitude. However we detected a positive correlation between copy number and the invasive distance from the origin area of the species, indicating that the variation of *mos1* activity is probably more related to the demographic history of *D. simulans* species. The colonization of

D. simulans to Europe is associated with a large loss of genetic diversity (Veuille *et al.* 2004), which could lead to the fixation of some TE insertions including *mos1*.

The *piRNA* pathway polymorphism and the *mos1* expression

In *Drosophila*, the *piARN* pathway is a key mechanism involved in the silencing of active TE families. The *piRNA*-mediated silencing is initiated by the transcription of an antisense RNA precursor from a few genomic *loci* called the *piRNA* clusters. This long precursor is then cleaved by proteins of the PIWI family into small RNAs 23-29 nt long (*piRNA*). Among the three PIWI proteins present in *Drosophila* genomes, *Piwi* and *Aub* are able to load antisense guide *piRNAs* that mediate TE mRNA recognition and decay. After this primary step is achieved, the third PIWI member *Ago3* can form a complex with *Aub* in order to initiate the production of sense *piRNAs* that will eventually feed the massive production of double stranded *piRNAs*. This “ping-pong” amplification cycle of TE-derived *piRNAs*, hence provides the necessary *piRNA* pool required to silence highly transcriptionally expressed TEs.

More than thirty genes are involved in the germline *piARN* regulatory machinery and mainly consist in gene-encoding proteins interacting with PIWI family members. Interestingly, coding sequences of *piRNA*-associated genes exhibit strong signature of positive selection between *Drosophila* species, suggesting that they rapidly adapt to their changing parasitic ecosystem (Kolaczowski *et al.* 2011; Lee & Langley 2012; Obbard *et al.* 2009b). Likewise, immunity genes involved in pathogen resistance, genes associated to *piRNA* pathway, are assumed to be involved in an evolutionary arms race against TE activity that results from intragenomic conflicts within the germline (Lee & Langley 2012). In *D. simulans*, transcription levels of seven *piRNA* genes were analyzed, and revealed a strong expression polymorphism between populations of different geographical origins (Fablet *et al.* 2014). In this work we observed a strong transcriptional polymorphism for both *Aub* and *Ago3* key regulatory genes consistent with an “immunity like” evolutionary scenario.

The different findings highlighted in the present work concerning *mos1* and the *piRNA* genes allow us to propose a scenario, summarized in Fig. 7. We observed that *mos1* activity increases in colonizing populations compared to the non-colonizing ones. It has been shown that the population size of colonizing populations is much smaller compared to the population of ancestral origin (Veuille *et al.* 2004), probably because only a few individuals successfully established new lineages “out of Africa”. Thus, natural selection is expected to be an effective force in non-colonizing populations with large population size, whereas drift prevails in colonizing populations with small population size. Considering that *mos1* is active in all populations studied (PMM>0), the low copy number in non-colonizing population may be explained by purifying selection efficient enough to remove even slightly deleterious *mos1* insertions. Inversely, the increase in copy number in colonizing populations could be an effect of drift having randomly maintained copies in the different genome lineages. Moreover, we observed that transcription of *mos1* could be connected to *piRNA* gene variations at different levels. On one hand, it can be noticed that transcription levels of *mos1* always display an opposite pattern to the transcriptional activity of *Ago3*. The high activity of *mos1* in testes matches the low expression of *Ago3* in the 19 populations, suggesting that *Ago3* is less efficient in the male germline. Inversely, on the other hand, the reduced *mos1* activity goes along with high *Ago3* expression in ovaries. However, in non-colonizing populations, whatever the sex, no correlation is observed between *mos1* and *Ago3*. The only significant correlation is found in ovaries of colonizing populations, in which *Ago3* is globally more transcribed, and *mos1* more sensitive to its variation. Thus it is probably under stronger control after its reactivation in a colonizing context.

The *piRNA* regulation has been recently integrated into models of TE dynamics (Lee &

Langley 2012; Lu & Clark 2010). It has been shown that host resistance to TEs is easily acquired if one copy jumps into a *piRNA* cluster (Khurana *et al.* 2011). Such an insertion can rapidly be fixed because of its selective advantage to silence transcription of the other active copies (Lu & Clark 2010). This has been exemplified by the high frequency of *P* element insertion at the 1A locus (Ronsseray *et al.* 1991) conferring resistance to the strong *P* element activation in natural populations. In parallel, it has been shown that in the absence of (or weak) *piRNA* control of TE transcription, new TE insertions into euchromatic regions are subject to strong negative selection, leading to their population removal or low frequencies maintenance (Lu & Clark 2010). Consistent with our observations, *mos1* insertions into *piRNA* clusters are probably more frequent in colonizing populations where *mos1* activity has been reactivated, and less frequent in non-colonizing populations where purifying selection continuously purge the genome from new *mos1* deleterious insertions.

Relaxed control of TE activity in testes

mos1 activity is several orders of magnitude higher in testes than in ovaries. On the contrary, we observed that *piRNA* gene expression is enriched in ovaries, a conserved pattern in all populations of *D.simulans*. Such large sex differences are usually observed for sexually dimorphic traits. The conserved pattern of female biased expression of *piRNA* genes indicates that these genes have evolved through female-specific selection pressures (Ellegren & Parsch 2007). It has been shown that the *piRNA* pathway is essential for both female germline maintenance and progeny development (Aravin & Bourc'h's 2008; Molaro *et al.* 2014; Vagin *et al.* 2006). Indeed, a main difference between the two germlines is the epigenetic transmission of a *piRNA* pool through the mothers but not the fathers. The female-biased expression pattern of *piRNA* genes observed here is thus consistent with high maternal investment to ensure the progeny optimum fitness. By contrast, these selective pressures seem relaxed in the male germline. A weak or absent *piRNA* regulation in testes can result in a male-biased TE activity, as observed for *mos1*, and potentially participate in the male-biased evolution of genomes.

A male-biased mutation rate is frequently observed in many groups of species (Connallon & Clark 2010; Connallon & Knowles 2005; Hurst & Ellegren 1998; Parsch & Ellegren 2013). This can be explained by the difference in the number of germ cell divisions, which is higher in males than in females (Sayres & Makova 2011). However, this is probably not the only explanation since many other replication-independent mechanisms were shown to be involved (Hurst & Ellegren 1998; Sayres & Makova 2011). Here, we observed high levels of *mos1* activity in testes of all *D. simulans* populations. The case of *mos1* is not unique in *Drosophila*, and several other TE families are over-expressed in the male germline (Haoudi *et al.* 1997; Kogan *et al.* 2003; Lankenau *et al.* 1994; Pasyukova *et al.* 1997). In addition, *LINEs* and *SINEs*, the most active Class of TEs in mammals, display male-specific activities both in terms of expression and genome distribution (Jurka *et al.* 2002; Nellaker *et al.* 2012). If the TE testis-specific activity is a general phenomenon, it could largely contribute to the faster rate of evolution driven by the male germline. However this question is still poorly investigated and further global approaches are needed.

Another phenomenon, called the “out of testes” hypothesis, was recently highlighted in several works. The authors showed that a majority of new genes emerged with a testis-specific expression pattern (Heinen *et al.* 2009; Kaessmann 2010; Levine *et al.* 2006; Light *et al.* 2014; Paulding *et al.* 2003; Ponce & Hartl 2006; She *et al.* 2004; Wang *et al.* 2000). Here again, the enhanced activity of TEs in testes could largely participate in this phenomenon by generating novel ORFs or transcription-promoting regions with testis-specific expression. For instance, many primate *de novo* genes expressed in testes consist in TE-derived exons (Toll-Riera *et al.* 2009). Many TE insertions close to genes were shown to be involved in the specific

recruitment of new transcription factor binding sites with testes specific expression (Ruiz-Orera *et al.* 2015). Numerous *de novo* genes are retrotransposition of mRNAs and non-coding RNAs, mediated by the reverse transcriptase of active TE families (Kaessmann 2010). In addition to these well-described genetic events, the amplification of TEs can lead to non-homologous ectopic recombination, and eventually give birth to new chimerical genes (Kaessmann 2010). By being more permissive for TEs, the male germline thus appears as a niche for molecular and gene tinkering, thus enlarging the selective spectrum and somehow promoting adaptability.

Acknowledgements

This work was supported by the Institut Diversité, Ecologie et Evolution du Vivant (IDEEV) and the APEGE program of the CNRS in environmental genomics for their financial support. We thank Malcolm Eden for review of the text.

Data accessibility

All informations are available in the manuscript and the supplemental data

Author contributions

BSL, AHV and PC designed research; BSL and IC performed research; BSL, IC, AHV and PC analyzed data; BSL wrote the manuscript with input of all authors.

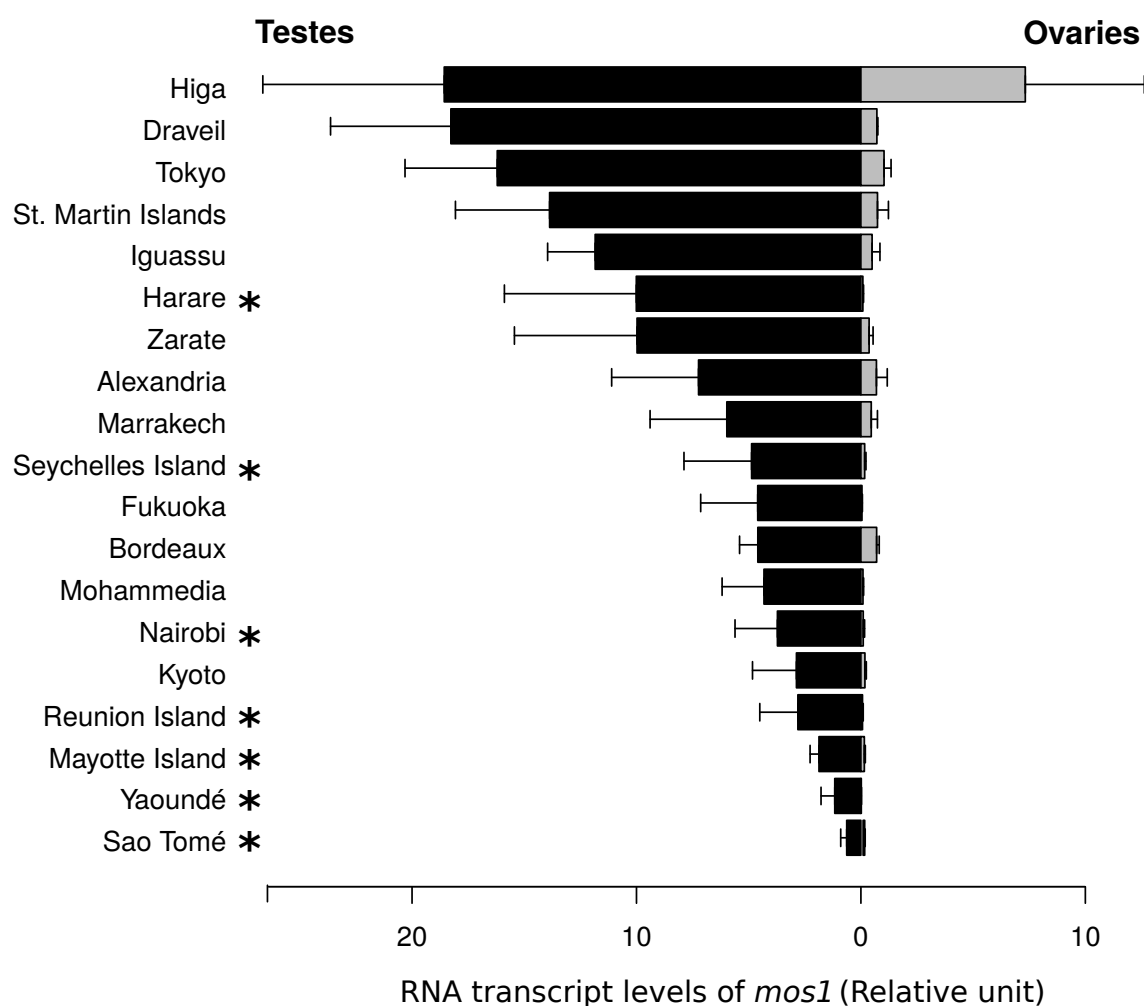


Figure 1: *mos1* expression in gonads (Black bars = testes, Grey bars = ovaries) of individuals from 19 natural populations. The reference is the expression *mos1* in ovaries of Marrakech. The natural populations are: Sao Tomé; Yaoundé (Cameroon); Mayotte; Reunion Island; Kyoto, Tokyo, Higa and Fukuoka (Japan); Nairobi (Kenya); Seychelles and St Martin Islands, Mohammedia and Marrakech (Morocco); Bordeaux and Draveil (France); Alexandria (Egypt); Harare (Zimbabwe); Zarate (Argentina); Iguassu (Brazil). Light bars = standard error; * = Sub-Saharan populations.

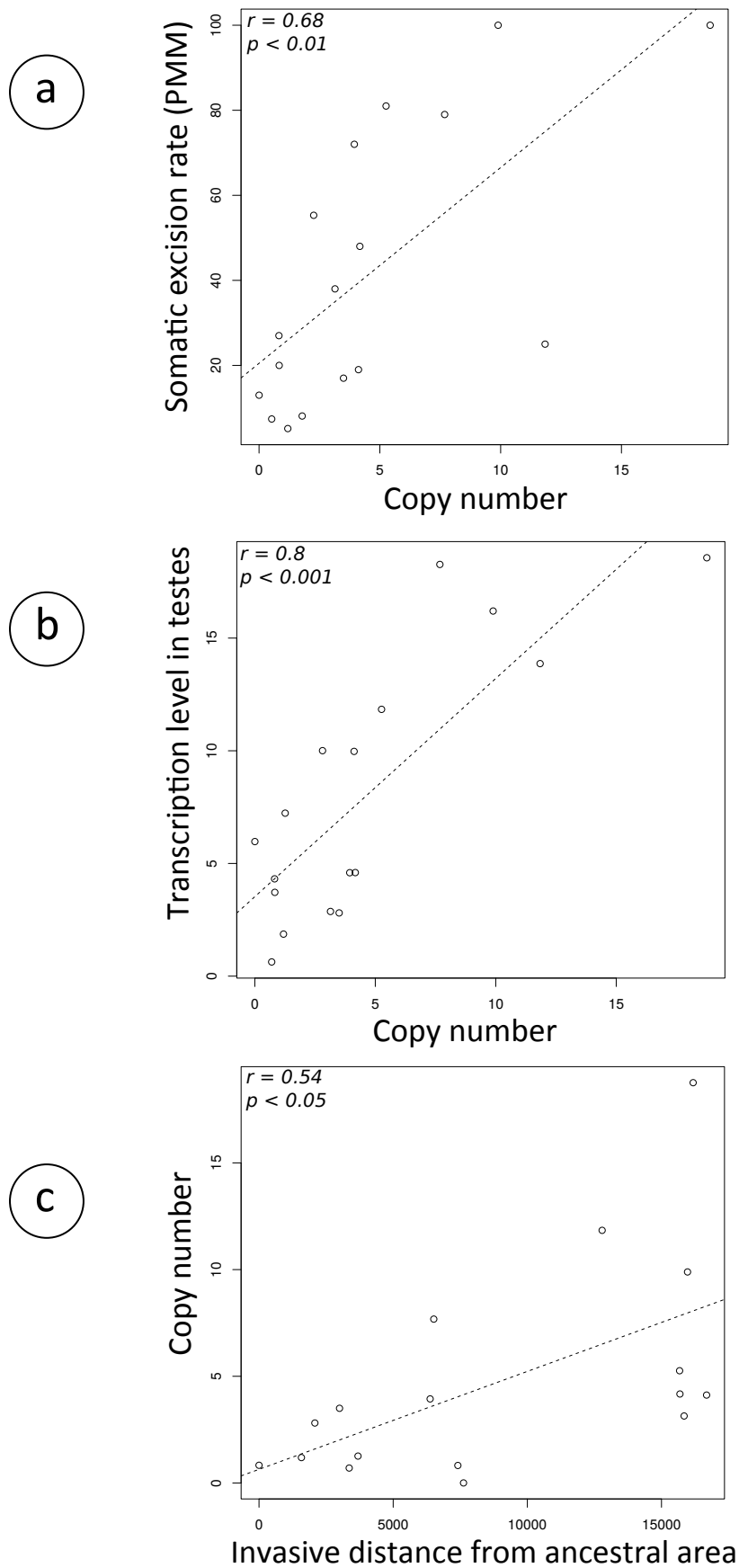


Figure 2: (a) Relationships between copy number and somatic excision rate; (b) copy number and transcription level in testes; (c) copy number and invasive distance from ancestral area (in km). All correlations are calculated using the Pearson coefficient.

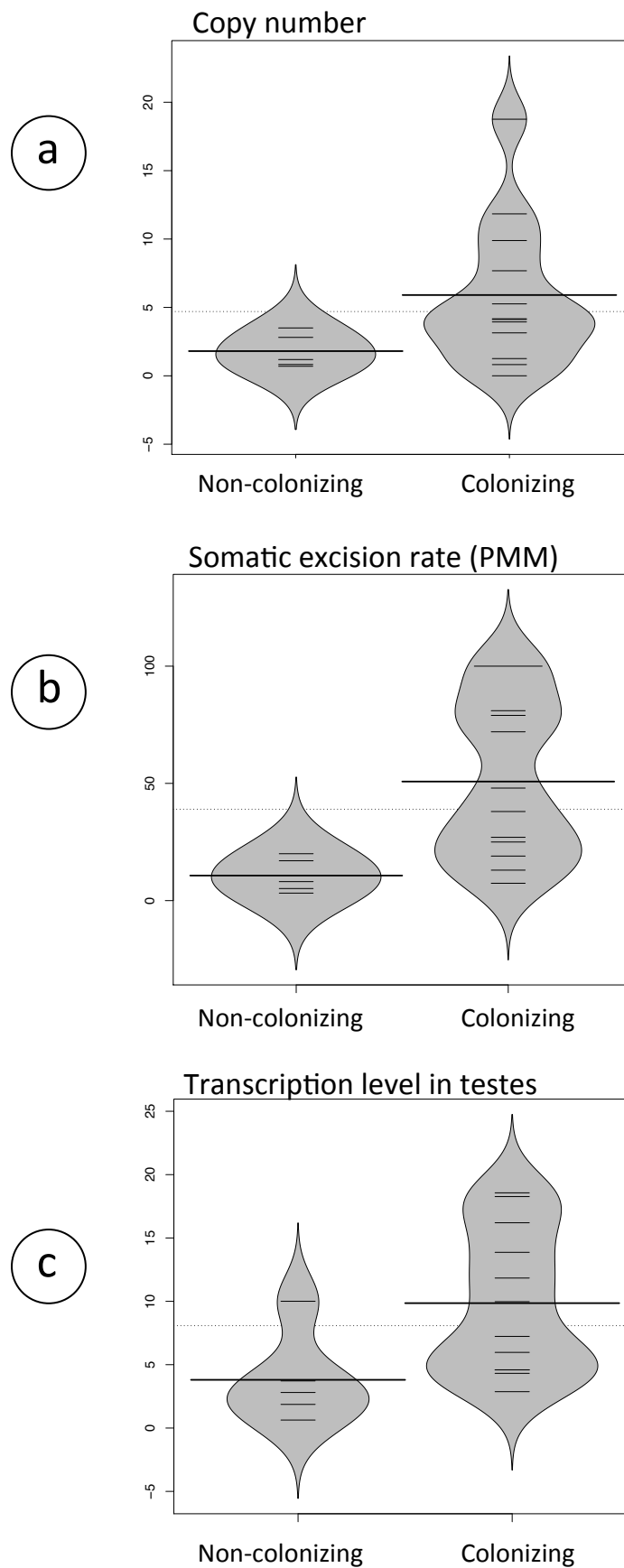


Figure 3: Violin plots (from R) showing the difference of (a) copy number, (b) somatic excision rate (PMM) and (c) transcription level of *mos1* in testes. Expression levels and variability of *mos1* are always higher in colonizing populations than in sub-Saharan ones.

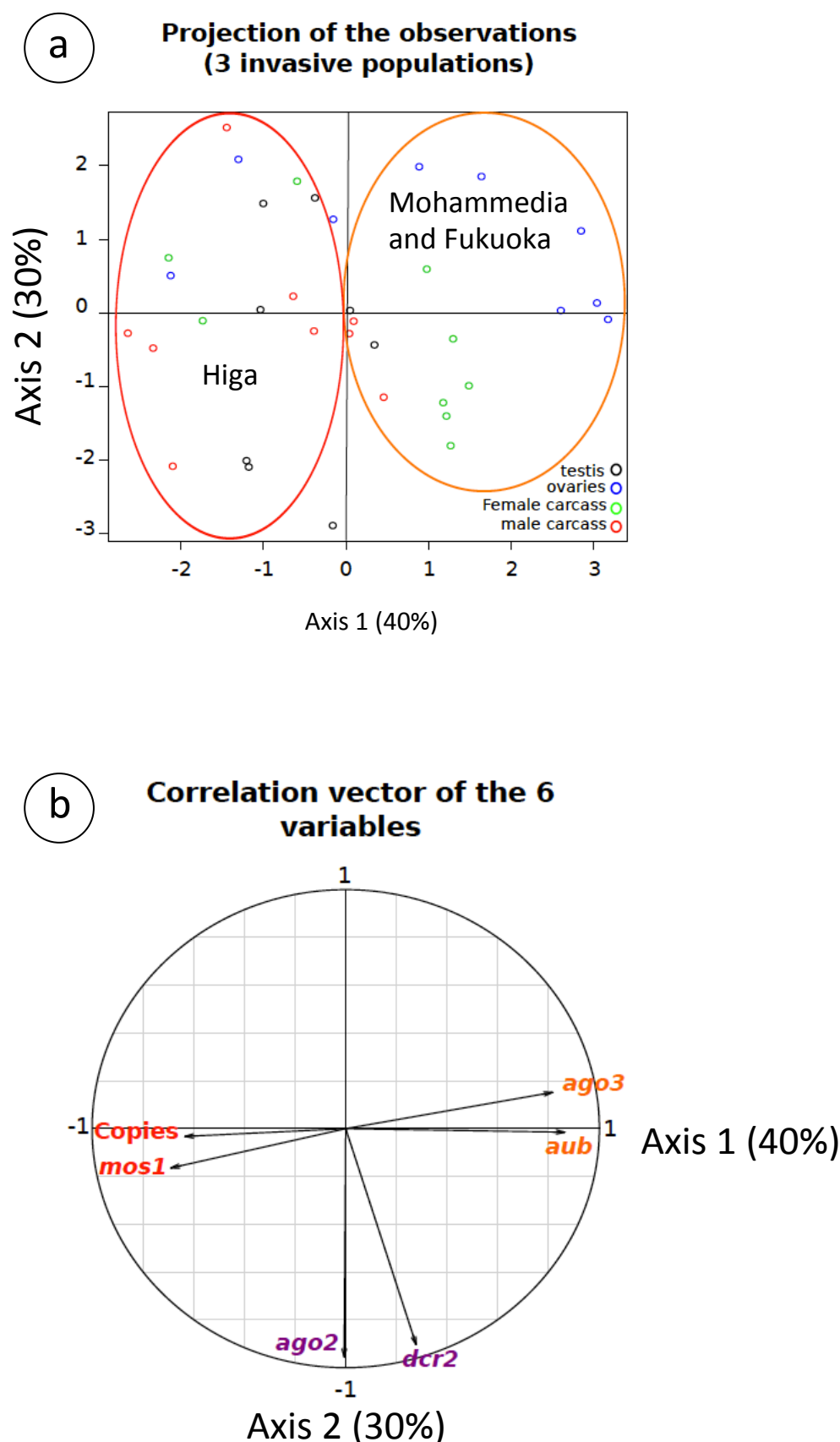


Figure 4: (a) Projection of the three populations (Higa, Fukuoka and Mohammedia) onto the first two axes of a Principal Component Analysis (PCA). The PCA was done using the expression levels (estimated by RT-qPCR) of *Ago3*, *Ago2*, *Dcr2*, *Aub* and *mos1* and the copy number as variables. Axis 1 represents 40% of the total variance and axis 2, 30%. Each point represents an individual. (b) Correlations between axes and variables showing that axis 1 is mainly explained on one hand by variations of *Ago3* and *Aub* expressions (positive correlations), and by variations of *mos1* expression and of copy number, on the other hand.

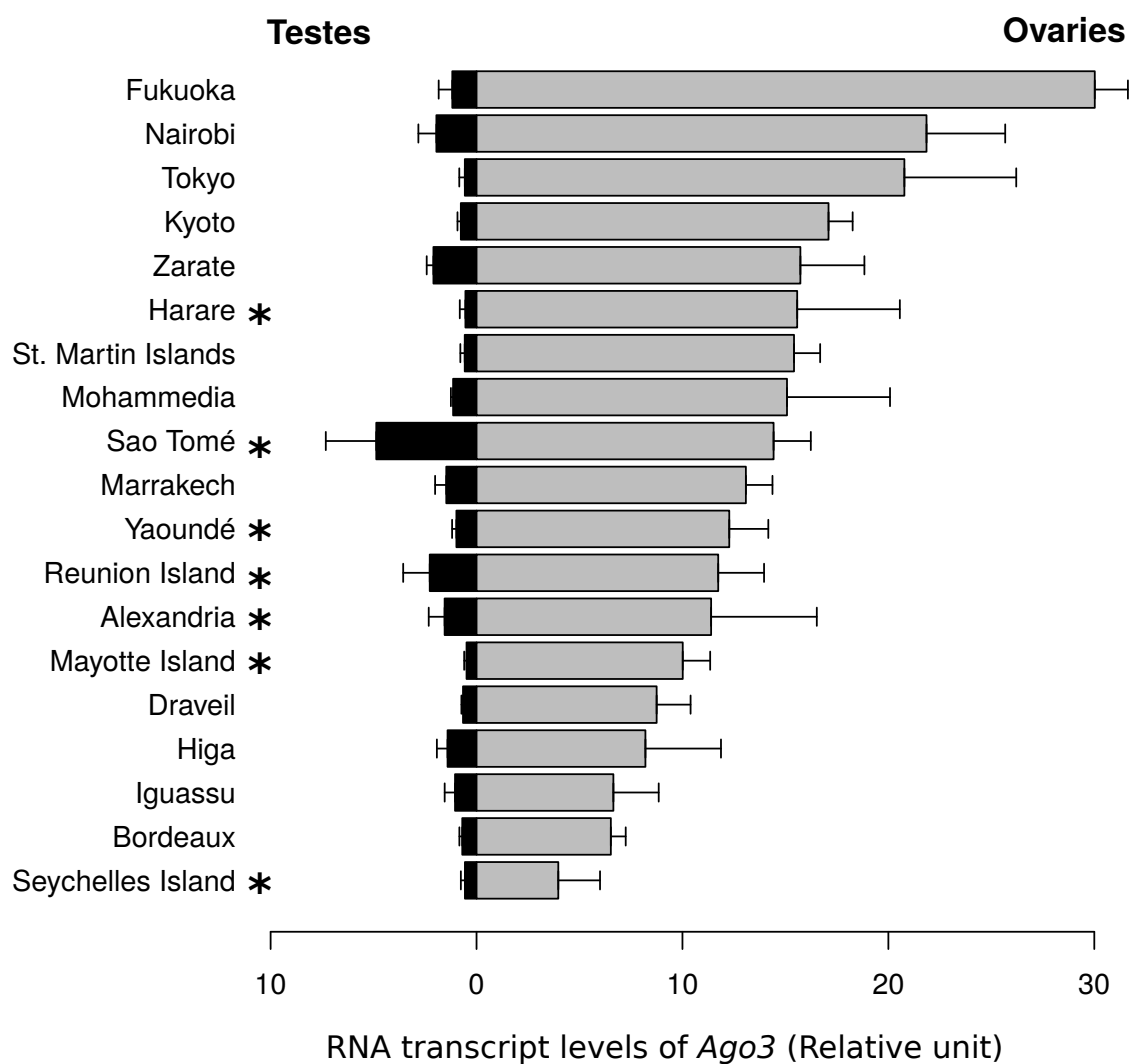
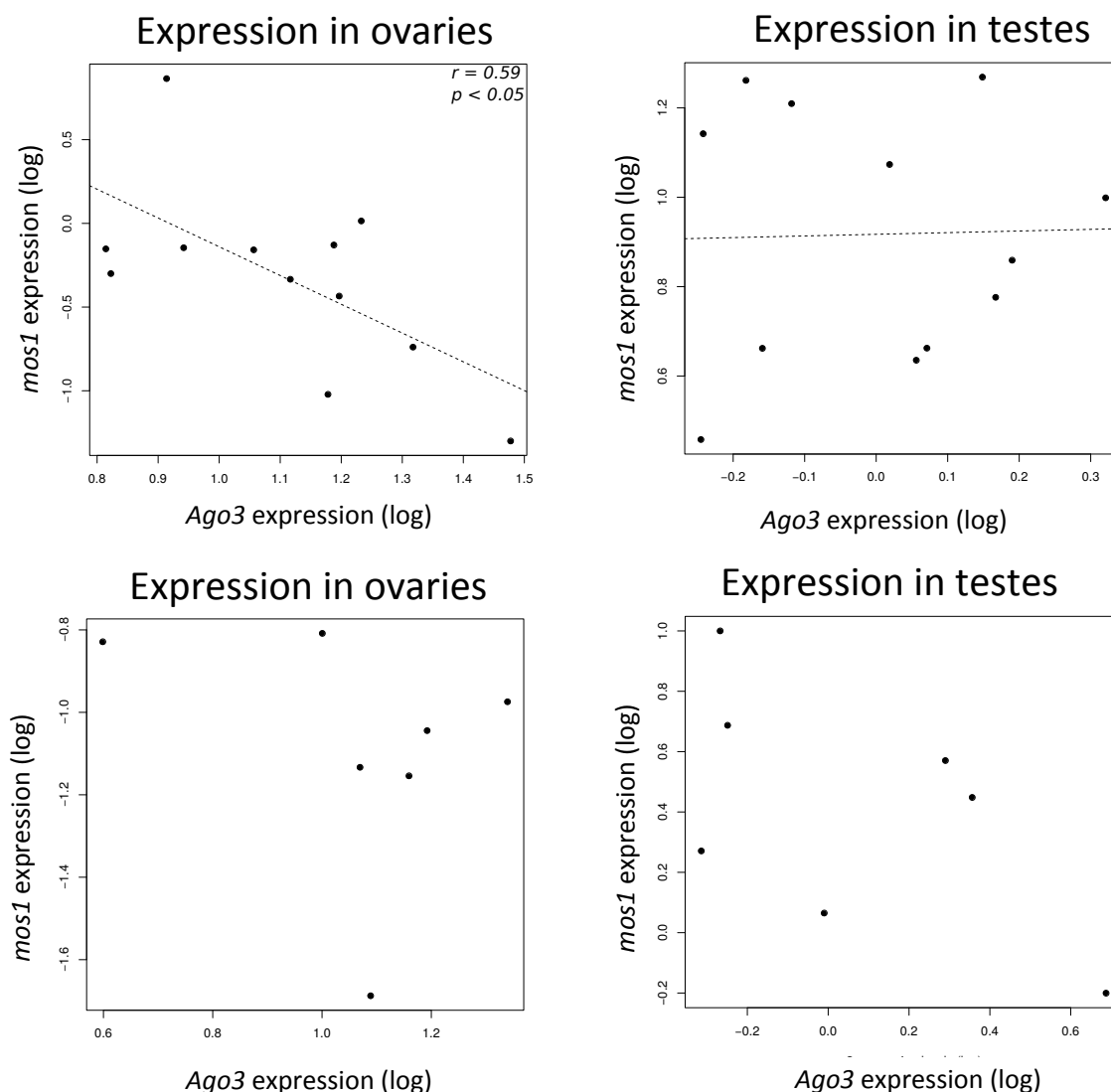


Figure 5: *Ago3* transcriptional activity in gonads of individuals sampled in 19 natural populations. Black bars = testes, Grey bars = ovaries. The reference is the *Ago3* expression in testes of Mohammedia. The natural populations are the same as those used in Figure 1.

Colonizing populations



Non-colonizing populations

Figure 6: Correlations between *mos1* and *Ago3* expressions in ovaries and testes of African (Sub-Saharan) and in colonizing populations. Correlations are calculated using the Pearson coefficient. Only the correlation in ovaries of colonizing populations is significant.

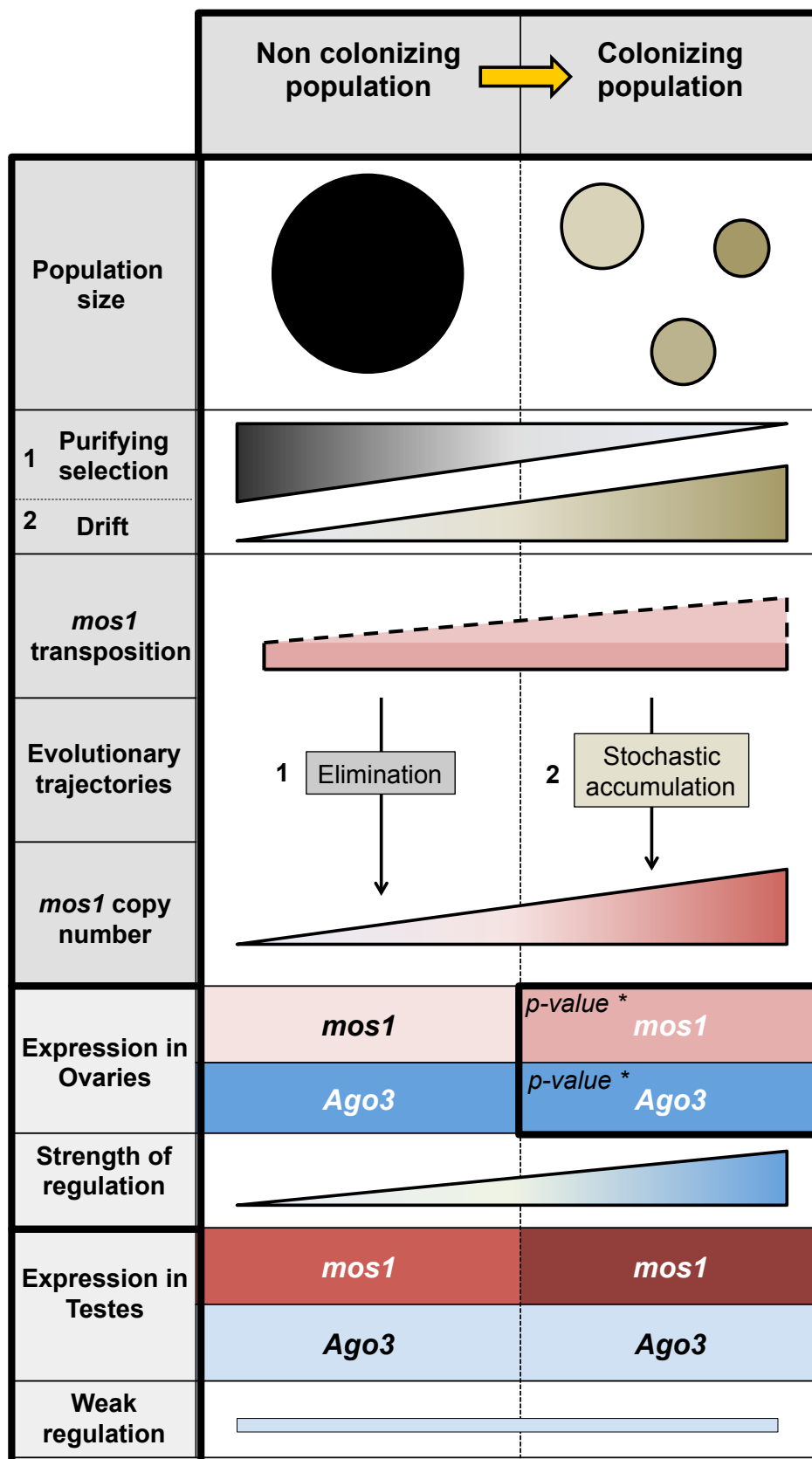


Figure 7: Summary of the evolutionary scenario proposed. Comparison of non-colonizing and colonizing populations is done according to several populational and evolutionary forces that are suspected to act on the *mos1* activity. The black circle represents the large effective population size of non-colonizing populations and grey circles illustrate the small population size of colonizing populations. Purifying selection (black triangle) and drift (brown triangle) are symbolized respectively by a decreasing vs increasing gradient according to population size. The *mos1* transposition and copy number (red triangle) represent the positive correlations observed in Fig. 2 and Fig. 3. The *mos1* (red rectangular) and *Ago3* (blue rectangular) expressions are illustrated by a qualitative colored panel (the color intensity is related to the relative expression level – light red = lower expression than deep red). Intensity of *pi*RNA regulation in ovaries is symbolized by an increasing blue gradient.

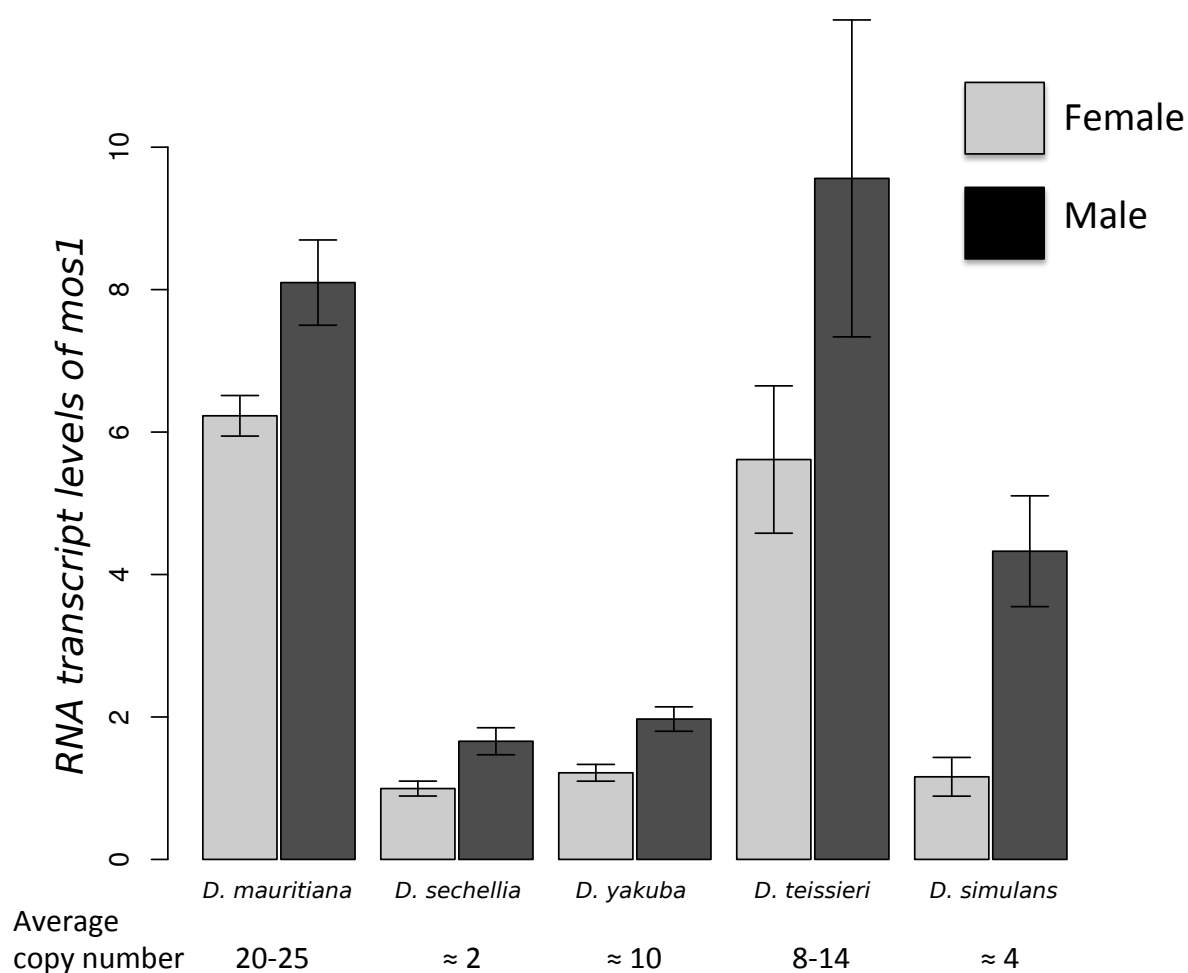


Figure S1: Difference males – females of transcript level for 5 species of the melanogaster subgroup. mau = *D. mauritiana*, sec = *D. sechellia*, yak = *D. yakuba*, tei = *D. teissieri*, sim = *D. simulans* (natural population from Fukuoka – Japan). The transcript levels are estimated from a sample of 20 individuals (males or females). Vertical bars represent the standard error. Copy numbers are extracted from literature as following: mau = Jacobson *et al.* 1986; sec = Capy *et al.* (1991); yak = Capy *et al.* (1992); tei = Brunet *et al.* (1996); sim = present work (pop = Fukuoka).

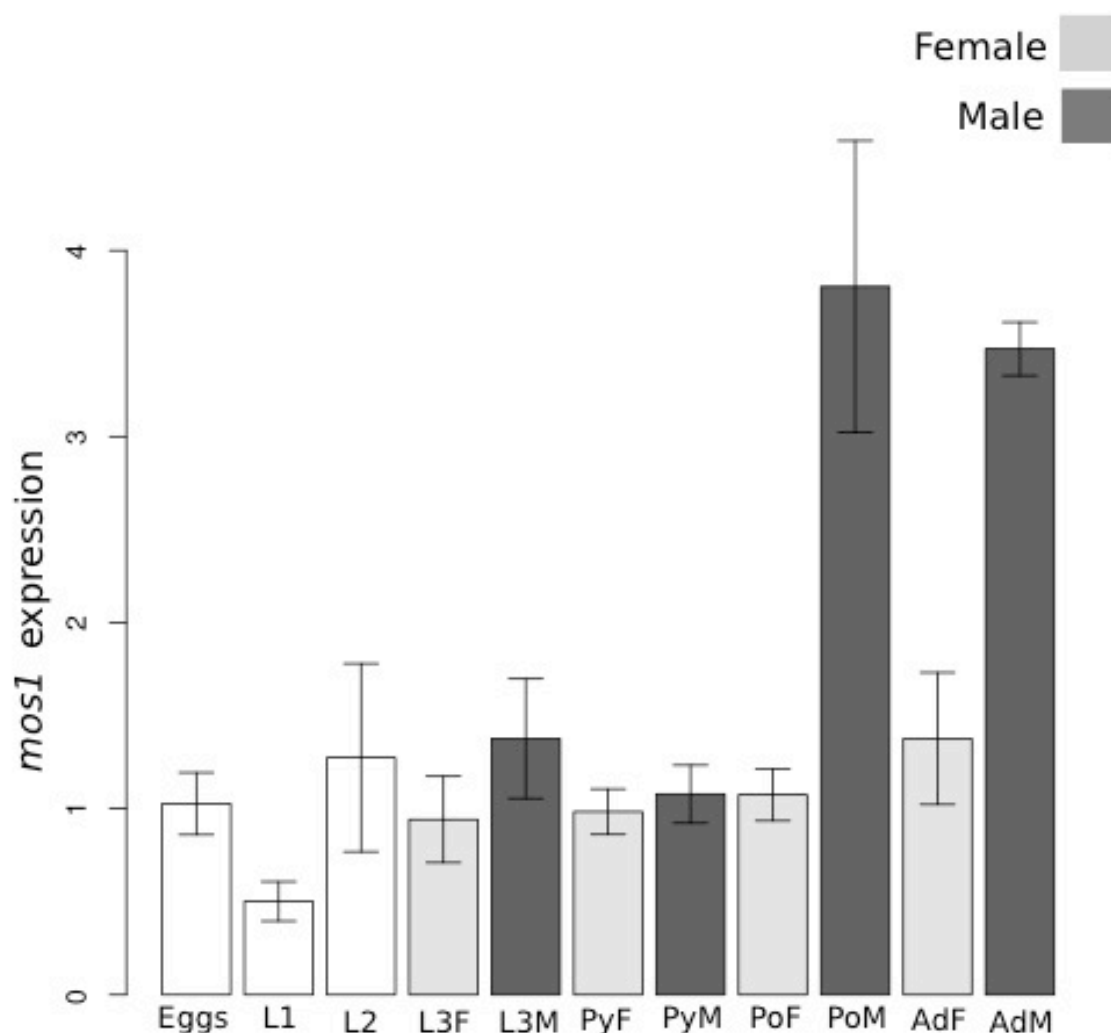


Figure S2: Male-Female differences of *mos1* expression during the development from eggs to adults in the natural population of Fukuoka (Japan). L1 = 1st-instar larvae, L2 = 2nd-instar larvae, L3 = 3rd-instar larvae, PY = Young pupae, PO = Old pupae, Ad = Adult, F = Female, M = Male. Individuals were sexed from L3 to adult. Visual determinations were confirmed by PCR on the sex-linked gene *vermillion*. Vertical bars represent the standard error.

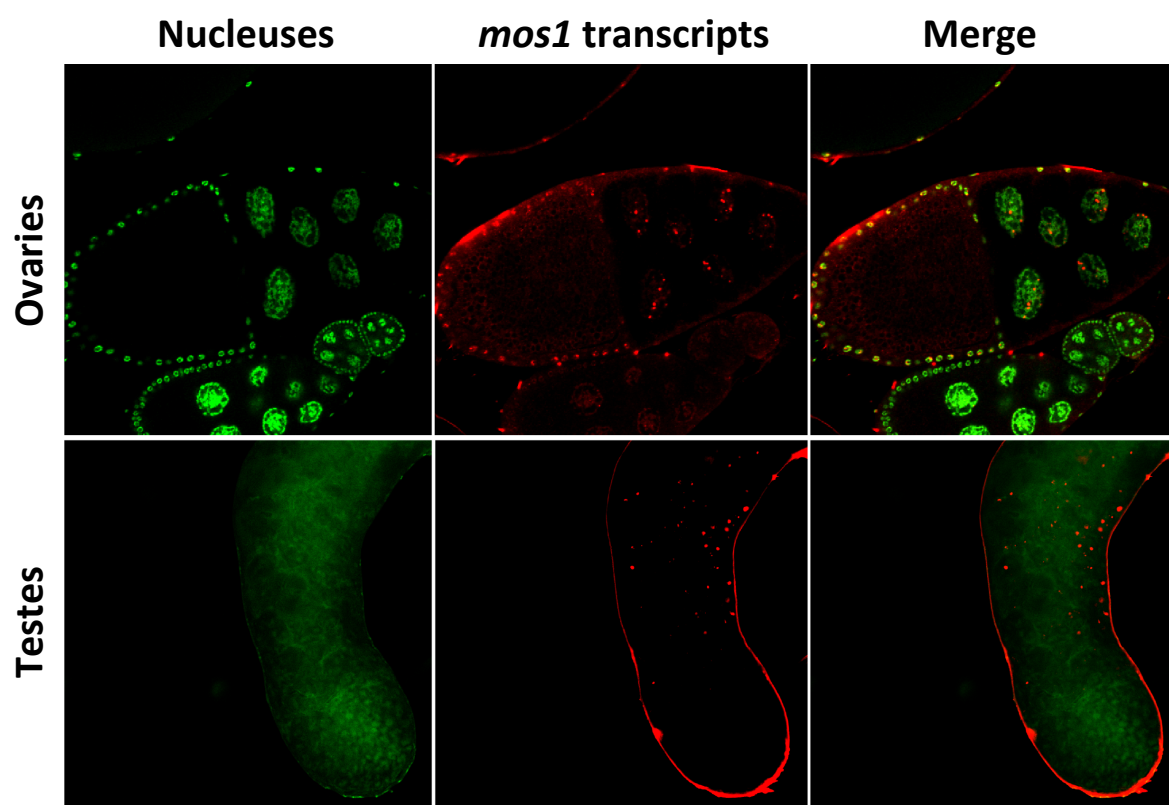


Figure S3: *In situ* hybridization of *mos1* on ovaries and testes on individuals originated from Higa (Japan). Nucleuses are in green and *mos1* senses transcripts are in red.

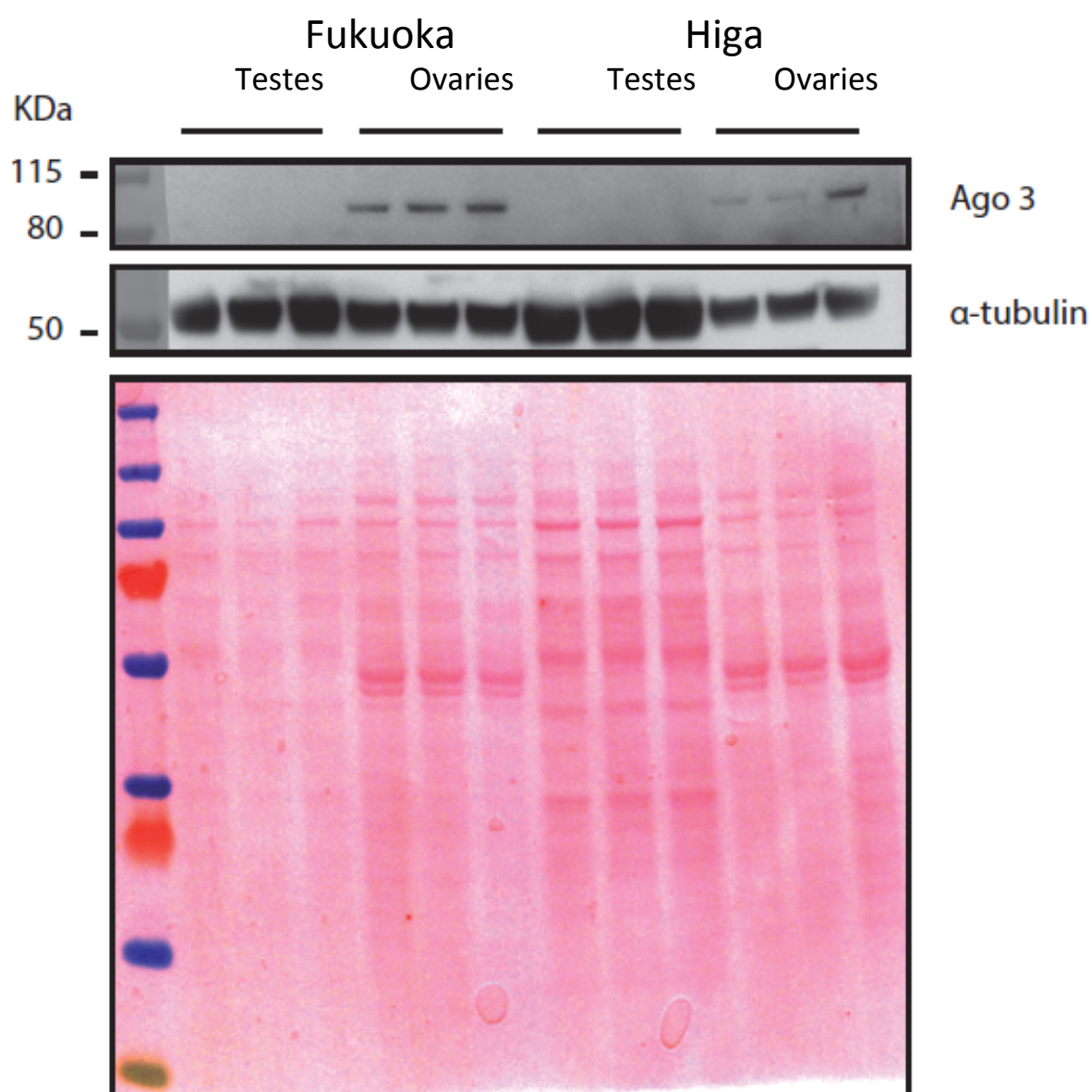


Figure S4: Western blot on ovaries and testes of two Japanese populations (Fukuoka and Higa). The antibody used was directed against the *Ago3* protein.

Table S1: List of primers used in the qPCR experiments

Primers for RNA quantifications of *mos1* and *piRNA* genes

<i>mos1</i> _Forward	GTGAACGGTGGTTTCAACG
<i>mos1</i> _Reverse	AGCGATTGGAACTGCTTGT
<i>ago3</i> _Forward	TTTTTCCTCCCATTGACTCG
<i>ago3</i> _Reverse	CCGTCAAATGTTTTGGTTCC
<i>ago2</i> _Forward	TGCAGTTCTTCATGGTCAGC
<i>ago2</i> _Reverse	CATGTGGCACAGGTTGTAGG
<i>aubergine</i> _Forward	GGGCCATGAATGAACATAACC
<i>aubergine</i> _Reverse	TGTCAAGCTCGATGTTCCAG
<i>dicer2</i> _Forward	TCTGCTCCATCAGTTTCACG
<i>dicer2</i> _Reverse	TTGATCCGCATCCAGTATGA

Primers used for normalization***Development*** (Fig.supp2)

RP49_Foward	CCAGTCGGATCGATATGCTAA
RP49_Reverse	GTTTCGATCCGTAACCGATGT
Light_Foward	CCGATTCCAAAGCTCACATT
Light_Reverse	TTGACAAAACACTGCCTTCG
RPL17q2_Foward	CCCTCCTTTTTCGTTTTTCGTT
RPL17q2_Reverse	GTGTTGTCGGCACAGTTCAT

Testes and Ovaries comparison (Fig.1 and Fig. 5)

His4r-F	AAGCGTCATCGTAAGGTGCT
His4r-R	CCGCGAGTTTCCTCGTAAAT
RpS29-F	ACGGTCTGATCCGCAAGTAT
RpS29-R	CTTCTTGAAGCCAATGTCGT

Species(Fig.supp1)

RP49 66F	CCAGTCGGATCGATATGCTAA
RP49 186R	GTTTCGATCCGTAACCGATGT
RPL11_F	CGATCCCTCCATCGGTATCT
RPL11_R	AACCACTTCATGGCATCCTC
RPL17q2-1	CCCTCCTTTTTCGTTTTTCGTT
RPL17q2-2	GTGTTGTCGGCACAGTTCAT

References

- Akkouche A, Rebollo R, Burlet N, *et al.* (2012) *tirant*, a newly discovered active endogenous retrovirus in *Drosophila simulans*. *Journal of Virology* **86**, 3675-3681.
- Aminetzach YT, Macpherson JM, Petrov DA (2005) Pesticide resistance *via* transposition-mediated adaptive gene truncation in *Drosophila*. *Science* **309**, 764-767.
- Andersen CL, Jensen JL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**, 5245-5250.
- Aravin AA, Bourc'his D (2008) Small RNA guides for *de novo* DNA methylation in mammalian germ cells. *Genes & Development* **22**, 970-975.
- Arnault C, Dufournel I (1994) Genome and stresses: reactions against aggressions, behavior of transposable elements. *Genetica* **93**, 149-160.
- Arthur AL, Weeks AR, Sgro CM (2008) Investigating latitudinal clines for life history and stress resistance traits in *Drosophila simulans* from eastern Australia. *Journal of Evolutionary Biology* **21**, 1470-1479.
- Begun DJ, Aquadro CF (1995) Molecular variation at the *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *Drosophila simulans*. *Genetics* **140**, 1019-1032.
- Biemont C, Nardon C, Deceliere G, *et al.* (2003) Worldwide distribution of transposable element copy number in natural populations of *Drosophila simulans*. *Evolution* **57**, 159-167.
- Bock IR, Parsons PA (1981) Species of Australia and New Zealand. In: *The Genetics and Biology of Drosophila* (ed. Ashburner M CH, Thompson JN Jr), pp. 291-308. Academic Press, London.
- Bouletreau-Merle J, Fouillet P, Varaldi J (2003) Divergent strategies in low temperature environment for the sibling species *Drosophila melanogaster* and *D. simulans*: overwintering in extension border areas of France and comparison with African populations. *Evolutionary Ecology* **17**, 523-548.
- Boussy IA, Itoh M, Rand D, Woodruff RC (1998) Origin and decay of the *P* element-associated latitudinal cline in Australian *Drosophila melanogaster*. *Genetica* **104**, 45-57.
- Brennecke J, Aravin AA, Stark A, *et al.* (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089-1103.
- Brunet F, Godin F, Bazin C, David JR, Capy P (1996) The *mariner* transposable element in natural populations of *Drosophila teissieri*. *J. Mol. Evol.* **42**, 669-675.
- Capy P, Gasperi G, Biémont C, Bazin C (2000) Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**, 101-106.
- Capy P, Gibert P (2004) *Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different. *Genetica* **120**, 5-16.
- Capy P, Koga A, David JR, Hartl DL (1992) Sequence variation among active *mariner* elements in natural populations of *Drosophila simulans*. *Genetics* **130**, 499-506.

Capy P, Maruyama K, David JR, Hartl DL (1991) Insertion sites of the transposable element *mariner* are fixed in the genome of *Drosophila sechellia*. *J. Mol. Evol.* **33**, 450-456.

Capy P, Pla E, David JR (1994) Phenotypic and genetic variability of morphometrical traits in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. 2. within population variability. *Genetics Selection Evolution* **26**, 15-28.

Casacuberta E, Gonzalez J (2013) The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**, 1503-1517.

Chakir M, Chafik A, Moreteau B, Gibert P, David JR (2002) Male sterility thermal thresholds in *Drosophila*: *D. simulans* appears more cold-adapted than its sibling *D. melanogaster*. *Genetica* **114**, 195-205.

Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package - I : One-table methods. *R News* **4**, 5-10.

Chung H, Bogwitz MR, McCart C, *et al.* (2007) *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* **175**, 1071-1077.

Connallon T, Clark AG (2010) Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* **64**, 3417-3442.

Connallon T, Knowles LL (2005) Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics* **21**, 495-499.

Daborn PJ, Yen JL, Bogwitz MR, *et al.* (2002) A single *P450* allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253-2256.

David JR, Bocquet C (1975) Similarities and differences in latitudinal adaptation of two *Drosophila* sibling species. *Nature* **257**, 588-590.

David JR, Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. . *Trends in Genetics* **4**, 106-111.

Dean MD, Ballard JWO (2004) Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. *Molecular Phylogenetics and Evolution* **32**, 998-1009.

Ellegren H, Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics* **8**, 689-698.

Fablet M, Akkouche A, Braman V, Vieira C (2014) Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis. *Gene* **537**, 149-153.

Garcia Guerreiro MP, Chavez-Sandoval BE, Balanya J, Serra L, Fontdevila A (2008) Distribution of the transposable elements *bilbo* and *gypsy* in original and colonizing populations of

Drosophila subobscura. *Bmc Evolutionary Biology* **8**.

Giraud T, Capy P (1996) Somatic activity of the *mariner* transposable element in natural populations of *Drosophila simulans*. *Proc. Roy. Soc. Lond. (Biol.)* **263**, 1481-1486.

Gonzalez J, Karasov TL, Messer PW, Petrov DA (2010) Genome-Wide Patterns of Adaptation to Temperate Environments Associated with Transposable Elements in *Drosophila*. *Plos Genetics* **6**.

Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008) High Rate of Recent Transposable Element-Induced Adaptation in *Drosophila melanogaster*. *Plos Biology* **6**, 2109-2129.

Gonzalez J, Macpherson JM, Messer PW, Petrov DA (2009) Inferring the Strength of Selection in *Drosophila* under Complex Demographic Models. *Molecular Biology and Evolution* **26**, 513-526.

Gonzalez J, Petrov DA (2009) The adaptive role of transposable elements in the *Drosophila* genome. *Gene* **448**, 124-133.

Gunawardane LS, Saito K, Nishida KM, *et al.* (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587-1590.

Hagan CR, Sheffield RF, Rudin CM (2003) Human *Alu* element retrotransposition induced by genotoxic stress. *Nature Genetics* **35**, 219-220.

Haoudi A, Rachidi M, Kim MH, *et al.* (1997) Developmental expression analysis of the 1731 retrotransposon reveals an enhancement of *Gag-Pol* frameshifting in males of *Drosophila melanogaster*. *Gene* **196**, 83-93.

Heinen TJAJ, Staubach F, Haeming D, Tautz D (2009) Emergence of a new gene from an intergenic region. *Current Biology* **19**, 1527-1531.

Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* **129**, 133-147.

Hurst LD, Ellegren H (1998) Sex biases in the mutation rate. *Trends in Genetics* **14**, 446-452.

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314.

Irvin SD, Wetterstrand KA, Hutter CM, Aquadro CF (1998) Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: Evidence for founder effects in new world populations. *Genetics* **150**, 777-790.

Jacobson JW, Medhora MM, Hartl DL (1986) Molecular structure of a somatically unstable

element in *Drosophila*. *Proc. Natl. Acad. Sc. USA* **83**, 8684-8688.

Jurka J, Krnjajic M, Kapitonov VV, Stenger JE, Kokhanyy O (2002) Active *Alu* elements are passed primarily through paternal germlines. *Theoretical Population Biology* **61**, 519-530.

Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Research* **20**, 1313-1326.

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6603-6607.

Khurana JS, Wang J, Xu J, *et al.* (2011) Adaptation to *P* Element Transposon Invasion in *Drosophila melanogaster*. *Cell* **147**, 1551-1563.

Kidwell MG (2002) Transposable elements and the evolution of the genome size in eukaryotes. *Genetica* **115**, 49-63.

Kidwell MG, Lisch D (1997) Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci U S A* **94**, 7704-7711.

Kofler R, Nolte V, Schloetterer C (2015) Tempo and Mode of Transposable Element Activity in *Drosophila*. *Plos Genetics* **11**.

Kogan GL, Tulin AV, Aravin AA, *et al.* (2003) The *GATE* retrotransposon in *Drosophila melanogaster*: mobility in heterochromatin and aspects of its expression in germline tissues. *Molecular Genetics and Genomics* **269**, 234-242.

Kolaczowski B, Hupalo DN, Kern AD (2011) Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Molecular Biology and Evolution* **28**, 1033-1042.

Kooke R, Johannes F, Wardenaar R, *et al.* (2015) Epigenetic basis of morphological variation and phenotypic plasticity in *Arabidopsis thaliana*. *Plant Cell* **27**, 337-348.

Kopp A, Frank A, Fu J (2006) Historical biogeography of *Drosophila simulans* based on Y-chromosomal sequences. *Molecular Phylogenetics and Evolution* **38**, 355-362.

Labrador M, Seleme WD, Fontdevila A (1998) The evolutionary history of *Drosophila buzzatii*. XXXIV. The distribution of the retrotransposon *Oswaldo* in original and colonizing populations. *Molecular Biology and Evolution* **15**, 1532-1547.

Lachaise D, Cariou ML, David JR, *et al.* (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evolutionary Biology* **22**, 159-225.

Lachaise D, Silvain JF (2004) How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster* - *D. simulans* palaeogeographic riddle. *Genetica* **120**,

17-39.

Lankenau S, Corces VG, Lankenau DH (1994) The *Drosophila microopia* retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. *Molecular and Cellular Biology* **14**, 1764-1775.

Lee S-I, Kim N-S (2014) Transposable elements and genome size variations in plants. *Genomics & informatics* **12**, 87-97.

Lee YCG, Langley CH (2012) Long-Term and Short-Term Evolutionary Impacts of Transposable Elements on *Drosophila*. *Genetics* **192**, 1411-+.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ (2006) Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9935-9939.

Light S, Basile W, Elofsson A (2014) Orphans and new gene origination, a structural and evolutionary perspective. *Current Opinion in Structural Biology* **26**, 73-83.

Lu J, Clark AG (2010) Population dynamics of *PIWI*-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Research* **20**, 212-227.

Makarevitch I, Waters AJ, West PT, *et al.* (2015) Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. *Plos Genetics* **11**.

McClintock B (1984) The significance of responses of the genome to challenge. *Science* **226**, 792-801.

Medhora MM, MacPeck AH, Hartl DL (1988) Excision of the *Drosophila* transposable element *mariner*: identification and characterization of the *mos* factor. *EMBO J.* **7**, 2185-2189.

Molaro A, Falciatori I, Hodges E, *et al.* (2014) Two waves of *de novo* methylation during mouse germ cell development. *Genes & Development* **28**, 1544-1549.

Morris CA, Benson E, White-Cooper H (2009) Determination of gene expression patterns using *in situ* hybridization to *Drosophila* testes. *Nature Protocols* **4**, 1807-1819.

Mourier T, Nielsen LP, Hansen AJ, Willerslev E (2014) Transposable elements in cancer as a by-product of stress-induced evolvability. *Frontiers in genetics* **5**, 156-156.

Nardon C, Deceliere G, Loevenbruck C, *et al.* (2005) Is genome size influenced by colonization of new environments in dipteran species? *Molecular Ecology* **14**, 869-878.

Nellaker C, Keane TM, Yalcin B, *et al.* (2012) The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biology* **13**.

Nolte V, Schlotterer C (2008) African *Drosophila melanogaster* and *D. simulans* populations have similar levels of sequence variability, suggesting comparable effective population sizes. *Genetics* **178**, 405-412.

Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM (2009a) The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical Transactions of the Royal Society B-Biological Sciences* **364**, 99-115.

Obbard DJ, Welch JJ, Kim K-W, Jiggins FM (2009b) Quantifying adaptive evolution in the *Drosophila* immune system. *Plos Genetics* **5**.

Parsch J, Ellegren H (2013) The evolutionary causes and consequences of sex-biased gene expression. *Nature Reviews Genetics* **14**, 83-87.

Pasyukova E, Nuzhdin S, Li W, Flavell AJ (1997) Germ line transposition of the *copia* retrotransposon in *Drosophila melanogaster* is restricted to males by tissue-specific control of *copia* RNA levels. *Molecular & General Genetics* **255**, 115-124.

Paulding CA, Ruvolo M, Haber DA (2003) The Tre2 (USP6) oncogene is a hominoid-specific gene. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2507-2511.

Pelisson A, Payen-Groschene G, Terzian C, Bucheton A (2007) Restrictive *flamenco* alleles are maintained in *Drosophila melanogaster* population cages, despite the absence of their endogenous *gypsy* retroviral targets. *Molecular Biology and Evolution* **24**, 498-504.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**, 509-515.

Picot S, Wallau GL, Loreto ELS, *et al.* (2008) The *mariner* transposable element in natural populations of *Drosophila simulans*. *Heredity* **101**, 53-59.

Ponce R, Hartl DL (2006) The evolution of the novel *Sdic* gene cluster in *Drosophila melanogaster*. *Gene* **376**, 174-183.

Ronsseray S, Lehmann M, Anxolabéhère D (1991) The maternally inherited regulation of *P* elements in *Drosophila melanogaster* can be elicited by two *P* copies at cytological site 1A on the X chromosome. *Genetics* **129**, 501-512.

Rouault JD, Casse N, Chenais B, *et al.* (2009) Automatic classification within families of transposable elements: Application to the *mariner* Family. *Gene* **448**, 227-232.

Rozhkov NV, Aravin AA, Zelentsova ES, *et al.* (2010) Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *Rna-a Publication of the Rna Society* **16**, 1634-1645.

Ruiz-Orera J, Hernandez-Rodriguez J, Chiva C, *et al.* (2015) Origins of *de novo* genes in human

and chimpanzee. *eprint arXiv:1507.07744*.

Russell AL, Woodruff RC (1999) The genetics and evolution of the *mariner* transposable element in *Drosophila simulans*: worldwide distribution and experimental population dynamics. *Genetica* **105**, 149-164.

Saito K, Siomi MC (2010) Small RNA-mediated quiescence of transposable elements in animals. *Developmental Cell* **19**, 687-697.

Sayres MAW, Makova KD (2011) Genome analyses substantiate male mutation bias in many species. *Bioessays* **33**, 938-945.

Schlenke TA, Begun DJ (2004) Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1626-1631.

Schoeßl G, Schloetterer C (2006) Microsatellite variation and differentiation in African and non-African populations of *Drosophila simulans*. *Molecular Ecology* **15**, 3895-3905.

Senti K-A, Brennecke J (2010) The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends in Genetics* **26**, 499-509.

She XW, Horvath JE, Jiang ZS, *et al.* (2004) The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857-864.

Singh RS, Choudhary M, David JR (1987) Contrasting patterns of geographic variation in the cosmopolitan sibling species *Drosophila melanogaster* and *Drosophila simulans*. *Biochemical Genetics* **25**, 27-40.

Sinzelle L, Izsvak Z, Ivics Z (2009) Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cellular and Molecular Life Sciences* **66**, 1073-1093.

Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology* **12**, 246-258.

Toll-Riera M, Castelo R, Bellora N, Mar Alba M (2009) Evolution of primate orphan proteins. *Biochemical Society Transactions* **37**, 778-782.

Vagin VV, Hannon GJ, Aravin AA (2009) Arginine methylation as a molecular signature of the Piwi small RNA pathway. *Cell Cycle* **8**, 4003-4004.

Vagin VV, Sigova A, Li C, *et al.* (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320-324.

Vandesompele J, De Preter K, Pattyn F, *et al.* (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, RESEARCH0034-RESEARCH0034.

Veuille M, Baudry E, Cobb M, Derome N, Gravot E (2004) Historicity and the population genetics of *Drosophila melanogaster* and *D. simulans*. *Genetica* **120**, 61-70.

Vieira C, Biemont C (2004) Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* **120**, 115-123.

Vieira C, Lepetit D, Dumont S, Biemont C (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Molec. Biol. Evol.* **16**, 1251--1255.

Wang W, Zhang J, Alvarez C, Llopart A, Long M (2000) The origin of the *Jingwei* gene and the complex modular structure of its parental gene, *yellow emperor*, in *Drosophila melanogaster*. *Molecular Biology and Evolution* **17**, 1294-1301.

Watanabe TK, Kawanishi M (1976) Colonization of *Drosophila simulans* in Japan. *Proceedings of the Japan Academy* **52**, 191-194.

Weeks AR, McKechnie SW, Hoffmann AA (2002) Dissecting adaptive clinal variation: markers, inversions and size/stress associations in *Drosophila melanogaster* from a central field population. *Ecology Letters* **5**, 756-763.

Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17600-17601.

Wisotzkey RG, Felger I, Hunt JA (1997) Biogeographic analysis of the *Uhu* and *Loa* elements in the Hawaiian *Drosophila*. *Chromosoma* **106**, 465--477.

Article n°3

***piRNA* and transposon dynamics in *Drosophila* : a female story**

Article n°3

*pi*RNA and transposon dynamics in *Drosophila* : a female story

Bastien SAINT-LEANDRE, Quentin HELLEU, Pierre CAPY, Jonathan FILEE

Introduction

Transposable elements (TEs) are almost ubiquitous genomic parasites characterized by their ability to replicate faster than any other genetic elements in the genomes. Repeatedly, TEs have been shown to contribute to genetic innovation (Bourque, 2009) and genome restructuring (Craig, 2002). However, TEs evolutionary trajectories largely depend on negative selective pressures acting against deleterious insertions (Dolgin and Charlesworth, 2008; Le Rouzic and Deceliere, 2005; Petrov et al., 2003). Indeed, TEs constitute strong mutagenic agents that can cause disruption of functional sequences, silencing of neighbor genes or favoring ectopic recombination events (Langley et al., 1988; Montgomery et al., 1987; Petrov et al., 2003).

In metazoan species germlines, TE families that exhibit strong activity will favor a vertical transmission through their amplification in gametes. However, strong TE activity in the germline is detrimental to the host fitness, causing severe deleterious phenotype (like gonadal dysgenesis and sterility) (Kidwell et al., 1977). It has been demonstrated that several species have independently acquired regulatory mechanisms as a response to TEs activity in germline. These regulations are based on interfering RNAs (Houwing et al., 2007; Kawaoka et al., 2009; Lau et al., 2006; Robine et al., 2009). Interestingly, studies in *Drosophila* were crucial in demonstrating a role of *pi*RNA mediated silencing in protecting the genome against their harmful effects into the female germline. Indeed, it has been shown in *Drosophila melanogaster* genome, that a discrete number of *loci*, called the *pi*RNA clusters, are specialized in the production of *pi*RNAs (Brennecke et al., 2007). From these *loci*, that represent less than 3% of the genome length, hundreds of thousands different *pi*RNAs are produced solely in the germline. Interestingly, most *pi*RNAs produced in the germlines are derived from TEs (up to 90% for some clusters) (Brennecke et al., 2007).

*pi*RNA are 23-29 nt RNAs found in complex with PIWI family proteins, playing crucial roles in transcriptional and post-transcriptional silencing of TEs. Antisense TE transcripts derived from the clusters, are first processed into *pi*RNAs serving as guides to target TE sense mRNAs. Then, the PIWI/*pi*RNA duplex initiates the TE transcript recognition and the mRNAs decay are mediated by PIWI proteins “slicing” activity. Depending on the nature of the *pi*RNA clusters and the PIWI-interacting proteins, the population of sense *pi*RNA can eventually feed a secondary amplification process called the ping-pong amplification loop. This process leads to a massive production of *pi*RNAs against a specific subset of active TE families (Brennecke et al., 2007; Gunawardane et al., 2007; Mohn et al., 2014). Thus, the *pi*RNA pathway has been proposed to serve as a genome defense mechanism against new invading TE families: when a TE fall into a *pi*RNA cluster, new guides RNAs are produced in order to silence active members of the same TE family.

Genetic studies of *D. melanogaster* ovaries have revealed that TEs are epigenetically silenced through the maternal transmission of PIWI proteins loaded with *pi*RNAs that are deposited into the oocytes. In the absence of this maternal *pool* of *pi*RNAs, inherited TEs become strongly activated in the progeny germline (Khurana et al., 2011). Indeed, transposons present in the male genome and absent in the female genome are

thus not silenced in the germline of the hybrid progeny. This can give rise to a termed hybrid dysgenesis syndrome associated with strong adult sterility, male recombination and gonadal atrophy (Bucheton et al., 1976; Kidwell et al., 1977). The two examples described so far in *D. melanogaster* are the *P-M* and the *I-R* systems respectively for the DNA transposon *P* element and the non-LTR retrotransposon *I* element (Brennecke et al., 2008; Chambeyron et al., 2008; Grentzinger et al., 2012; Khurana et al., 2011). In these two cases, the syndrome highlights the strong relationship between the *pi*RNA pathway, as a genomic response against invading TE, and the germline activity of a unique TE family. However, a recent study has revealed that crosses between *D. virilis* strains lead to a complex case of dysgenesis that arise due to the global derepression of multiple active TE families in the germlines (Erwin et al., 2015). In this case, the rapid divergence of TE families between close related strains gave rise to divergent *pi*RNA outcome involving strong hybrid incompatibilities.

Drosophila melanogaster and *Drosophila simulans* are two close species that diverged 3 to 5 million years ago (Kelleher et al., 2012). When crossed together, hybrid females of *D. melanogaster* and *D. simulans* exhibit abnormal germline phenotypes and sterility associated with the global derepression of TE families in ovaries. In this case, the TE derepression cannot be explained by the pool of *pi*RNA maternally inherited because all TE families are concerned (Kelleher et al., 2012). Moreover, interspecific hybrids phenocopy *pi*RNA pathway mutants lead to *i*) a global *pi*RNA production deficiency *ii*) an aberrant *pi*RNA cluster activity and *iii*) a loss of the *pi*RNA secondary amplification cycle (Kelleher et al., 2012). The authors came to the conclusion that, due to the important TE landscape divergence between the two species, *pi*RNA effector proteins have been constrained by diversifying selective pressures. The *pi*RNA pathway associated proteins thus became rapidly divergent and unable to play their functional role in interspecific hybrids. Independent studies have shown that *pi*RNA associated proteins display recurrent signatures of positive selection and further belong to the faster evolving proteins in *Drosophila* genomes (Kolaczowski et al., 2011; Langley et al., 2012; Obbard et al., 2009a). According to the rapid changes at the DNA level, it has been shown in *D. simulans* that genes involved in the *pi*RNA pathway exhibit high levels of transcriptional polymorphism between populations of distinct geographical origins (Fablet et al., 2014). It is assumed that such an evolutionary pattern is driven by an arms race resulting from the conflicting interactions between *pi*RNA regulatory genes and TEs activity (Langley et al., 2012).

Genomic comparison between the two species has revealed that the TE content and landscape are dramatically different (Kofler et al., 2015; Lerat et al., 2011; Vieira et al., 1999). Most TE insertions segregate at low frequencies in populations of both species (Kofler et al., 2015), leading to large variations between natural populations (Vieira et al., 1999). Several evidences suggest that *D. melanogaster* genome have undergone recent transpositional bursts of many TE families (Bergman and Bensasson, 2007; Bowen and McDonald, 2001; Kofler et al., 2015). Furthermore, the sequence similarity of TEs from divergent lineages, suggests that the *D. melanogaster* genome has been repeatedly invaded by horizontally transferred novel TEs (Bartolome et al., 2009; Gilbert et al., 2010; Sanchez-Gracia et al., 2005). Moreover, it seems that TE insertions in *D. simulans* genome consist mainly in a large number of degraded copies whereas *D. melanogaster* genome contains a large number of full length copies in accordance with the recent activation of TEs in *D. melanogaster* (Lerat et al., 2011).

It has been suggested that the *pi*RNA pathway can control new bursts of TE activity at different levels (Blumenstiel, 2011; Kelleher and Barbash, 2013; Song et al., 2014). First, in *Drosophila* species, the codons biased of the genes implicated in the *pi*RNA pathway are positively correlated with the genome TE load. This observation suggests that the transcriptional efficiency of the *pi*RNA machinery has increased in species massively invaded by TEs (Blumenstiel, 2011). Second, recent invading TE families that reach high copy number in the genome correspond to the most abundant class of TE-derived *pi*RNAs. This result indicates that enhanced *pi*RNA production can stop (or slow down) the TEs expansion (Kelleher and Barbash, 2013). However, comparative studies of *pi*RNA repertoires between populations of both *D. simulans* and *D. melanogaster* lead to intriguing results. Indeed, despite the fact that these two species have a very different TE

contents in their genomes, each of them display low variation levels of expression of TE derived *pi*RNAs between populations (Akkouche et al., 2013; Song et al., 2014). This observation raises interesting questions regarding the exact relationship between TE activity and their regulation by the *pi*RNA pathway in these two sibling species. Moreover, since the discovery of *pi*RNAs in *D. melanogaster* testes involved in the *Stellate* mRNA silencing (a gene for which the expression causes male sterility and leads to the formation of star-shaped crystals in testes) (Aravin and Tuschl, 2005), few studies have taken attention to TE-*pi*RNAs interactions in the male germline (Nagao et al., 2010; Nishida et al., 2007).

In the present work, we first compare transcriptomes of both males and females germlines in several populations of *D. simulans* and *D. melanogaster*. We found that most of the TE expressed in gonads display significant sex-biased transcription patterns. In addition, we characterized the TE-*pi*RNA relationship in the germlines using deep sequencing libraries of both testes and ovaries. Interestingly, our results show that most of the TE-derived *pi*RNAs are produced in ovaries whereas TE-derived *pi*RNAs production seems absent in testes. Comparison between *D. melanogaster* and *D. simulans* revealed different TE-*pi*RNA relationships. Indeed, *D. melanogaster* ovary *pi*RNAs preferentially match with TE families overexpressed in testes that show signatures of recent bursts of transposition. Inversely, in *D. simulans*, *pi*RNAs preferentially match fragmented and inactivated copies found to be overexpressed in ovaries. We suggest that these important differences in term of TE regulation between these two sibling species are linked to their different TE invasion history:

- Ancient invasion in *D. simulans* followed by a current control of the old TE families
- Ongoing expansion in *D. melanogaster* and progressive implementation of the *pi*RNA machinery.

The results presented here, highlights the complex TE/*pi*RNA dynamic that occurred all along the female germline during the 5 millions years of *D. melanogaster* and *D. simulans* separated histories.

Results- PART-1 *D. melanogaster*

1. Most of the TE families expressed in the germlines have a male-biased expression in *D. melanogaster* genome

Using a custom library of Drosophila elements consensus derived from Repbase, we first analysed the genome TE content in *D. melanogaster* (Supplementary Figure 1). The total repeat fraction of the *D. melanogaster* genome counts for 23% of the genome, including satellites and simple repeats. Most of these repeats are composed by TEs that counts for more than 18% of the genome. TEs that belong to the LTR retrotransposons constitute the most represented class of TE (12%) followed by non-LTR retrotransposons (5%) and DNA transposons (1%).

In order to understand the relationship between the TE load and their germline activity, we have sequenced testes and ovaries transcriptomes of two populations of *D. melanogaster*. Among the (471) TE families expressed (see methods), we found a difference of expression between ovaries and testes in 187 different TE families. Interestingly, the relative proportion of TEs showing sex-biased transcription is closed to the global genomic TE content, with a dominance of LTR retrotransposons (Figure 1A and Supplementary Figure 1). We restrict our genomic analysis to sex-biased TEs in order to address the question of their evolutionary tempo. Size distribution of all TE families found to be sex-biased in *D. melanogaster* was analysed. The total copy number of sex-biased TE families was plot according to their size in the reference genome (Figure 1B). We found the majority of the copies (69,5%) to be larger than 1kb (approximately the size of a *mariner* DNA transposon). Indeed, the majority of the copies found in the *D. melanogaster* genome correspond to an elevated number of canonical full-length elements that further poorly diverge between them. These results indicate that most of the TE families highly expressed in the germlines result from recent amplification bursts in *D. melanogaster*.

In addition, the proportion of TE having a sex-biased expression increases according to their expression levels (normalized mean number of reads in both gonads of Gotheron and Zimbabwe *D. melanogaster* populations – Figure 1C and D). For instance, the elements displaying low expression levels (mean number of read < 10) display low proportions of sex-biased TEs (male-biased and female-biased). By contrast, most of the highly expressed TEs families have a sex-biased expression pattern. Strikingly, most of them were found to be overexpressed in testes. Indeed, in the Gotheron population, 80% of the TE families expressed over the range 100 are sex-biased: 60 % male-biased and 20 % female-biased. In the Zimbabwe population, at the same threshold, more than 60 % of TEs have a sex-biased pattern of expression (50% male-biased and 10% female-biased).

2. *pi*RNA mediated silencing of TE is effective in ovaries but not in testes

In order to better understand the nature of the TE regulation by the *pi*RNA pathway, we have sequenced two small RNA banks coming from testes and ovaries of *D. melanogaster*. We have mapped these banks and other public *D. melanogaster* small RNA banks on our TE consensus. In ovaries of *D. melanogaster*, TE-derived small RNAs are produced both in sens and antisens. The distribution length of sens RNAs (Figure 2A) is centered at 26 nucleotides (nt), as well as antisens RNA. Thus, the larger fraction of the small RNAs sequenced in ovaries corresponds to *pi*RNAs length (23 to 29 nt). Among these *pi*RNAs, a large fraction shows a typical ping-pong signature suggesting that they are involved in the secondary biogenesis process (Supplementary Figure 2A). We further mapped testes *pi*RNAs to the same TE consensus list. Surprisingly, we found that TE-derived *pi*RNAs are extremely rare compared to the situation in the ovaries (Figure 2B). The small RNA size distribution in testes does not display a high peak enrichment for 25-27 nt in both sens and antisens RNA, indicating that few of them seems involved in the secondary biogenesis. These results show that *pi*RNA mediated silencing mainly take place in the ovaries. In testes, *pi*RNAs are rare and do not derived from TEs, this would explain why the majority of TEs expressed in the germlines appears to be overexpressed in the testes. By opposition, in the ovaries, the abundance of *pi*RNAs targeting TEs strongly repressed the activity of these elements.

3. TEs families that have recently expanded are strongly silenced by the *pi*RNA machinery in the ovaries

We have previously shown that the majority of TEs expressed in the germlines have a testis-biased expression, probably because TE-mediated silencing is not efficient in testes. At this stage, an important question was to determine if the testis-specific overexpression pattern of most TE families is proportional to their degree of amplification in the host genome. From our *Drosophila* TE database, we designed a list of condition-specific TE families present in the genome of *D. melanogaster* (testis-overexpressed, ovary-overexpressed and not biased). From this list, we discriminated TEs into two classes according to their copy numbers (Figure 3A). Thus, we distinguished the class 1-20 clusters old and extinct families as well as young elements that have not yet expanded. Conversely, families that have recently expanded in the genome belong to class >20. RNA mapping to consensus families included, reads that mapped to small portion of complete size TE families, TE fragments from which size were inferior to 1kb have been discarded from the analysis (thus the 471 consensus families presented Figure 1 fall to 99 families here).

Among the total number of TE families, 65 belong to the lower class (1 to 20 copies) and 34 to the upper class (more than 20 copies). Interestingly, TEs that show no sex differences mainly belong to the lower class (27 TE family over 31) whereas TEs displaying a sex-biased expression accounts for a large number of highly amplified TEs (30 TE family over 68). Indeed, among the 34 families displaying a high degree of amplification, 30 display a sex-biased pattern of expression, mainly towards an overexpression in testes (20 out 30). This result suggests that sex-biased TEs mainly result from multiple and recent bursts of transposition.

In order to understand the associated host response by the *pi*RNAs, we mapped ovarian *pi*RNAs to our list of TE consensus according to their sex expression-biased (Figure 3B and C). For both conditions, the length distributions of small RNAs show that sense and antisense *pi*RNAs are produced, with an antisense peak at 26 nt. However, the total amount of *pi*RNAs derived from TEs overexpressed in testes (Figure 3B) is several orders of magnitude higher compared to those derived from ovary-biased TEs (Figure 3C). Among small RNAs of 23 to 29 nt length, *pi*RNAs showing the ping-pong signature is at 22% for TEs overexpressed in testes and less than 10% for TEs overexpressed in ovaries. These results show that the overexpression of recently active TE families in testes is the result of a strong silencing in ovaries, using *pi*RNAs involved in the secondary biogenesis. Moreover, the intensity of the *pi*RNA response is proportional to the TE family amplification level: higher is the copy number of a given family, stronger is the associated *pi*RNA response (Supplementary Figure 2A). Furthermore, the increase in *pi*RNA quantity goes along with higher levels of *pi*RNA involved in the ping-pong amplification cycle (Supplementary Figure 2A).

We then checked this global tendency by looking at each TE individually (Supplementary Figure 3A). This analysis confirmed the fact that, the observed correlation is not biased by few specific, highly amplified TE families displaying an extreme *pi*RNA response. These observations are thus consistent with the idea that the genome limits the expansion of highly active transposons by *pi*RNAs involved in the secondary amplification process. In addition, these data indicate that the genome response has been specifically enhanced by recent transposition bursts of certain TE families in *D. melanogaster*.

Results-PART-2 *D. simulans*

1. TEs expressed in the germlines display an equilibrate proportion of testis and ovary-biased TEs in *D. simulans*

The total repeat fraction of the *D. simulans* genome is strikingly less compared to *D. melanogaster* (Supplementary Figure 4). Only 8.5% of the genome length derives from repeated sequences. By contrast, only 4.5% of the *D. simulans* genome is composed of TEs, more than four times less compared to *D. melanogaster*. Interestingly, despite the strong TE content difference, the relative proportion of each class of TEs is quite similar in both species, with a dominance of LTR elements (Figure 4A).

We then analyzed the size distribution of each copy of the sex-biased TE families present in the *D. simulans* reference genome (Figure 4B). We found a minority of the copies (33.6%) to be longer than 1kb. The majority of the copies found in the *D. simulans* genome correspond to deleted elements that consist in fragmented pieces of canonical full-length TEs. These results show that sex-biased TEs that have been fixed in *D. simulans* mainly consist in inactive copies that are reminiscent of past TE invasions. Thus, these data revealed that TEs expressed in the germline of the two species underwent divergent TE evolutionary histories: sex-biased TEs of *D. melanogaster* result from large and recent waves of amplification (Figure 1B) whereas those of *D. simulans* are the result of ancient amplification waves and are now found as fragmented elements (Figure 4B). Compared to *D. melanogaster* populations, the proportion of testis versus ovary-biased TE in Fukuoka and Nairobi populations of *D. simulans* is more balanced, especially considering higher expression levels (Figure 4B and C). For instance, among TE families expressed over the range 300, 40% of TEs are overexpressed in ovaries and 40 % overexpressed in testes. However, likewise in *D. melanogaster*, highly expressed TE families in *D. simulans* display a strong tendency to be sex-biased.

In summary, both species have a similar TE diversity, but the *D. melanogaster* genome TE load is considerably higher than the TE content of the *D. simulans* genome. By itself, the TE excess of the *D. melanogaster* genome (25Mb) explains the genome size difference with *D. simulans* (175 Mb and 150 Mb

respectively). As recently suggested (Lerat et al., 2011), this difference could be explained by different evolutionary histories: recent TE invasion in *D. melanogaster* whereas ancient in *D. simulans*. Moreover, the TE evolutionary trajectory that differs between these two sibling species probably underlies differences for the gonad specific activity and the genome response provided through the *pi*RNA silencing machinery.

2. As in *D. melanogaster* TE *pi*RNA mediated silencing is solely effective in ovaries

We have sequenced two small RNA banks of *D. simulans* coming from testes and we have downloaded diverse ovaries RNA banks from public databases. In *D. simulans* ovaries, TE-derived small RNAs are produced both in sens and antisens. We identified a peak that range from 24 to 29 nt for both sens and antisens RNAs (Figure 5A). The larger fraction of the small RNAs sequenced in ovaries correspond to *pi*RNAs length, with a strong the ping-pong signature (Supplementary Figure 2B). As in *D. melanogaster*, the mapping of testes small RNAs to the same TE consensus list, shows that *pi*RNAs do not target TEs in testes (Figure 5B). The small RNA size distribution in testes does not show any peak enrichment, neither strong ping-pong signatures. As previously shown in *D. melanogaster*, these results show that *pi*RNA mediated silencing is involved solely the ovaries.

3. *D. simulans* ovarian *pi*RNAs target preferentially TEs overexpressed in ovaries

In *D. simulans* the majority of TEs expressed in the germlines are differentially expressed between male and female. An important question was to determine if a sex-specific expression is related to high degree of amplification, as we observed in *D. melanogaster* (Figure 3A). Among the total number of TEs, 74 belong to the lower class (1 to 20 copies) and 28 to the upper class (more than 20 copies). Interestingly, most of the highly amplified TEs belong to the sex-biased expressed TE families (26 out of 28) (Figure 6A). Compared to *D. melanogaster*, we first observed that there are more TE families displaying a low amplification range, consistent with the fact that most TE families in this genome are ancient and extinct. Second, if we focus on sex-biased TE families, we can notice that the majority of highly amplified TE families do not display an overexpression in the testes as observed in the *D. melanogaster* transcriptomes. Indeed, 17 out of 26 these TE families are overexpressed in the *D. simulans* ovaries.

We thus mapped *D. simulans* ovarian small RNAs on the sex-biased TE families in order to test if the *pi*RNA mediated response differs in a different genomic TE environment. The length distributions of small RNAs that have mapped on testis-biased and ovary-biased TEs, is presented in the Figure 6B and C. The two distributions show features of *pi*RNAs with an antisens peak ranging from 24 to 29 nt. Interestingly, the total amount of *pi*RNAs derived from TE overexpressed in ovaries (Figure 6C) is several orders of magnitude higher compared to those derived from TEs overexpressed in testes (Figure 6B). Also, among these *pi*RNAs, around 30% (Supplementary Figure 3B) show a significant ping-pong signature, which is significantly less, compared to ovary-biased TE derived *pi*RNAs (40%). Strikingly, the situation contrasts with the *D. melanogaster* ovarian *pi*RNAs that preferentially silence highly amplified TE families found to be overexpressed in testes. However, in *D. simulans*, the proportion of sexually differentially expressed TE is more equilibrate (Figure 4 C and D), consistent with the weak ovarian *pi*RNA response towards TE families overexpressed in testes. Such differences could also be explained by the higher degree of ovary-biased TE amplification, with the existence of a positive correlation between copy number and *pi*RNA abundance (Supplementary Figure 2B). Analysis of each TE family individually has confirmed that an increase in copy number is clearly associated to an enhanced production of double stranded *pi*RNAs (Supplementary Figure 3B). Indeed, the two species display the same kind of *pi*RNA responses: a massive production of *pi*RNAs by the ping-pong cycle, matching with the most amplified TE families in the genome. However, TE families within these species do not share similar

evolutionary features (Figure 1B and 4B): the young and active amplified TEs of *D. melanogaster* are preferentially targeted by secondary *pi*RNAs in ovaries (leading to an overexpression in testes). By contrast, the inactive fragmented copies of *D. simulans*, which did amplify in the past, are still under *pi*RNAs control. This, suggests that the regulation has persisted over time, due to the maintenance of fragmented copies inserted into *pi*RNA producing *loci*. Such *pi*RNA producing *loci* have probably been kept due to their advantage to repress other active members of the same family. Meanwhile, active TEs were progressively lost or maintained at low frequencies due to *pi*RNA transcriptional silencing.

Part-3 From current to past: differential TE regulation in *D. melanogaster* and *D. Simulans*

In both *D. melanogaster* and *D. simulans*, we observed that *pi*RNA regulation mainly occurs in ovaries. One expectation should be that, genes involved in the *pi*RNA pathway have been selected for, and thus exhibit pattern of female specific selection. In addition, a strong TE mobilization was shown to be associated to an enhanced *pi*RNA response suggesting that the *pi*RNA production is shaped by the tempo and the dynamic of the TE amplification. Thus, one would expect strong selective pressures on *pi*RNA genes in ovaries of *D. melanogaster*, whereas in *D. simulans* we would expect a relaxation of the selective constraints.

In order to test these assumptions, we have analysed transcriptional expression patterns of the 33 genes known to be essential to the germline *pi*RNA silencing mechanism (Handler et al., 2013) (Figure 7A and B). We have compared the expression of these genes between ovaries and testes from the transcriptomes of each population. In *D. melanogaster* (Figure 7A), most of the *pi*RNA regulatory genes (28 out of 33) are overexpressed in ovary, with the exception of *qin*, *papi* and *yb*, three *TUDOR* domains proteins that have a testis-biased expression pattern. In *D. simulans* (Figure 7B) we also observed that the majority of the *pi*RNA genes are overexpressed in ovaries. However, only 19 out of 33 of these genes are significantly overexpressed in ovaries, whereas 10 genes are overexpressed in testis. Among the 10 genes displaying a testis overexpression, we found three proteins (Cutoff, Deadlock, and Rhino) that have been described as keys initiators of the secondary *pi*RNA biogenesis in dual strand clusters (Mohn et al., 2014). Interestingly, the protein Deadlock exhibits the strongest expression fold change in testes compared to ovaries.

Several studies have shown that *pi*RNA genes evolved under diversifying selections (strong DN/DS biased), assuming that their evolution is more rapid than under a neutral rate of evolution in *drosophila* species. It has been proposed that sex-biased genes are maintained as a consequence of sexual antagonism and generally evolved at higher evolutionary rates (Parsch and Ellegren, 2013). In our case, most of the *pi*RNA genes were found to be ovary-biased in both species, which is consistent with the idea that *pi*RNA processing is under a female specific selection. The *pi*RNA mechanism involves the production of billions of *pi*RNAs targeting genome wide locations. This machinery thus appears to be a cost-effective mechanism that has been favored solely in the *Drosophila* female germlines. However, in *D. simulans*, fewer genes exhibit a female specific expression pattern and we found several key *pi*RNA genes less expressed in ovaries. These data are consistent with some relaxation of the selective pressures acting on the *pi*RNA machinery in *D. simulans*. This could be due to the fact that most amplified TE families are now degenerated and potentially constitute hundreds of self-regulatory sites that would have balanced the requirement of highly ovary expressed *pi*RNA genes.

Discussion

Here, we will first discuss the *D. melanogaster* and *D. simulans* differences concerning the tempo of TE activity over the past. Second we will discuss the features of the genome response mediated by the *piRNA* in relation with the TE activity. Finally we will address the evolutionary constraints and consequences of the strong sexual dimorphism observed both for TEs and genes associated with the *piRNA* regulatory pathway.

Divergent TE evolutionary history among *Drosophila* species

D. melanogaster and *D. simulans* are two close species that have diverged from 3-5 Mya. Here, we showed that their genomes display a very different TE content (Supplementary Figure1 and 4). These results are consistent with several previous studies about genome TE content variations between *D. melanogaster* and *D. simulans* (Lerat et al., 2011; Vieira et al., 2012; Vieira et al., 1999). Recently, a large-scale analysis from natural populations of both species has revealed that most the new TE insertions are due to the ongoing expansion of 58 TE families (Kofler et al., 2015). However, these recently invading TE families are predominantly characterized by extremely high levels of heterogeneity between populations and are thus, not present in the reference genomes (or not yet annotated). In our case, we only consider TE families that have been fixed in the reference genome and thus, it excludes low frequency TE lineages that are not yet established. For instance, the ongoing *P* element invasion of *D. simulans* (Kofler et al., 2015) is not present in the reference genome. However, our estimations of TE diversity (Figure 1 and 4) are similar to those determined by pool sequencing analyses (Kofler et al., 2015). Indeed, the most striking difference observed here, is the strong over-representation of deleted copies in *D. simulans* (Figure 1 and 4) compared to *D. melanogaster*. That last point indicates that the main species differences are related to the tempo of TEs activity. *D. melanogaster* is characterized by recent transposition bursts of numerous TE families and contrast with *D. simulans* where TEs content, mainly consist in fragmented and inactive elements. All together, these results indicate that bursts of TE activity were older in *D. simulans* compared to *D. melanogaster*.

In this study, from transcriptomic data of gonads from different populations, we also report that the TE transcriptional differences between populations are limited. By contrast, huge TE transcriptional differences exist between the male and the female germlines. We thus focused on TE families found to be differentially expressed in the germlines of both species. Interestingly, we found that the germline differences mainly concerned TE families with high levels of expression, a feature associated with high genomic copy number (Figure 3 and 6). One important feature that can modulate TE activity in the germline, is the *piRNA* pathway that specifically target sens TE mRNAs in order to transcriptionally, or post-transcriptionally, silence their expression. However, the relation between TE activity and the genome *piRNA* response is poorly understood and rarely implemented in models of TE dynamics.

TE activity and the *piRNA* genome response

We analysed the *piRNA* profiles derived from TE with different deep sequencing data from both *D. melanogaster* and *D. simulans* species. Analyses have been focused on ovarian TE-derived *piRNAs* in order to clarify the relationship between TE activity and the *piRNA* regulation. We observed that the majority of TE derived *piRNAs* match with TE families that have been successfully amplified in the genome (more than 40 copies) and share high degree of similitude (Figure 3 and 6). Interestingly, we also observed that this genomic response to TE invasion is predominantly achieved by secondary *piRNAs* that are derived from double strand *piRNA* clusters (as shown by the ping-pong signature, Supplementary Figure 2). These results are in accordance with a previous analysis on *D. melanogaster* that shows a positive correlation between *piRNA* abundance and signatures of recently amplified TEs (Kelleher and Barbash, 2013).

By contrast with well-established TE lineages, we found in both species, that TEs containing less than 5 copies in the genome (Supplementary Figure 2) exhibit strikingly different *piRNA* profiles. Indeed, Class 1-5 TEs are poorly targeted by *piRNAs* and are further not involved in *piRNA* secondary amplification cycle, suggesting that they are not yet fully targeted by the *piRNA* pathway. These observations also corroborate previous findings showing that TEs present in rare genomic locations do not correlate with *piRNA* abundance (Kelleher and Barbash, 2013). Interestingly, these results favor a model in which the *piRNA* silencing act as a post genomic response to the TE invasion. In this model, the genome is able to massively produced *piRNAs* against TEs only after their fixation at numerous loci in the genome.

One would expect that this scenario is shaped by natural selection that acts against the harmful effects of the most active TE families (Brennecke et al., 2007; Castaneda et al., 2011; Malone and Hannon, 2009). It is known that *piRNA* production is initiated in specific regions called the *piRNA* clusters. These clusters are composed of repeated sequences that are mainly TEs and their fragmented derivatives (Brennecke et al., 2007). Their genomic locations are conserved across *Drosophila* species indicating that natural selection have restricted the *piRNA* production to few dedicated *loci* (Malone and Hannon, 2009). Interestingly, it seems that the pool of TEs within a *piRNA* cluster, is subject to rapid turnovers suggesting that the genome frequently update its regulatory repertoire by new insertions of active TE family (Khurana et al., 2011; Malone and Hannon, 2009). In parallel, models of TE dynamics predict that since *piRNA* silencing is implemented, a new insertion into a *piRNA* cluster can rapidly reach fixation (Lu and Clark, 2010). Indeed, a new insertion into highly heterochromatic regions is *i)* not deleterious to the host and *ii)* eventually confers the host an advantage to silence other TEs. By opposition, euchromatic copies are often associated to deleterious effects and thus frequently removed by purifying selective forces (Gonzalez et al., 2008; Lu and Clark, 2010).

One speculation would be that in the later stages of the TE invasion, copies trapped into *piRNA* clusters progressively accumulate at a faster rate than euchromatic copies. This hypothesis is consistent with the positive correlation between copy number and *piRNA* abundance and the predominance of heterochromatic TE insertions in both species. In this case, the comparison between *D. melanogaster* and *D. simulans* is of particular interest because they display different tempo of TE activity. *D. simulans piRNAs* mainly target old TE families whereas *D. melanogaster piRNAs* mainly target younger ones (Supplementary Figures 2, 4 and 5). Interestingly, comparison between TE families amplified at same levels in the two species, systematically lead to a higher ping-pong signature in *D. simulans* (Supplementary Figure 2). These results support the hypothesis that ancient TE invaders have been accumulated in the *piRNA* producing *loci*. Also, these observations show that once established, the *piRNA* silencing persists until the complete loss of the ancient families. In the very late stages of invasion, while full length active element keep on decline or become extinct, reminiscent copies are still able to persist as small TE relics embedded within the *piRNA* clusters.

The TE *piRNA* regulatory machinery is a female-specific system

In both *D. melanogaster* and *D. simulans* species, we found that *piRNA* silencing is principally effective in ovaries. In testes of both species, we found that *piRNAs* constitute a weak fraction of our small RNA libraries and very few of them exhibit a ping-pong signature (Figures 2 and 5). In *D. melanogaster*, most of the TE families were found to be overexpressed in testes (Figure 1) because of their preferential silencing in ovaries (Figure 3). However, we did not find a similar pattern in *D. simulans* transcriptome, in which TEs do not show this strong bias toward a testes-overexpression (Figure 4). Interestingly, these observations fit well with the hypothesis of a differential mode of TE activity between the two species. Indeed, TE families that are overexpressed in the testes of *D. melanogaster*, correspond to massively amplified active TE lineages that are actively silenced in ovaries by the *piRNA* machinery. However, TEs overexpressed in *D. simulans* testes do not show features of recent activity and are thus not optimally target by *piRNAs* in ovaries (Figure 3). Interestingly, the main targets of *D. simulans* ovarian *piRNAs* correspond to families that have undergone old waves of amplification and now appearing as fragments of the canonical TEs (Figure 4), indicating that the TEs have

been actively silenced in the past and still produce *pi*RNAs.

However TE-derived *pi*RNA production in ovaries is not the only way by which, the genome adapt to its pool of endogenous parasites (Lee and Langley, 2012). The evolutionary arms race between the host and TEs also have direct consequences on *pi*RNA effector proteins rates of evolution (Aravin et al., 2007; Blumenstiel, 2011; Kidwell and Lisch, 2001; Lee and Langley, 2012; Siomi et al., 2008). Independent works support that *pi*RNA proteins belong to the faster evolving component of coding sequences in *Drosophila* genomes, and are further subject to recurrent adaptive mutations (Berry et al., 2009; Heger and Ponting, 2007; Kolaczowski et al., 2011; Lee and Langley, 2012; Obbard et al., 2009b; Pane et al., 2007).

From our transcriptomic datas, we compared pattern of expression of 33 genes known as essential to the *pi*RNA mediated silencing of TEs in the female germline of *D. melanogaster* (Handler et al., 2013). We found that the vast majority of these genes tend to be overexpressed in ovaries (Figure 7). It is usually expected that genes under sex-specific selection display a sex-biased expression (Ellegren and Parsch, 2007). In this sense, the *pi*RNA regulatory genes are under a strong female selection in both species. However, we have also noticed substantial differences between the two species. Only 3 genes show a testis-biased expression in *D. melanogaster* whereas 10 are overexpressed in *D. simulans* testes. The 3 genes overexpressed in *D. melanogaster* testes probably have male specific functions in addition to *pi*RNA regulation, as exemplify by the *yb* gene involved in the male germline stem cell renewal (Szakmary et al., 2009). These three genes were also found to be less expressed in *D. simulans* ovaries in addition to 7 other genes that also displayed ovary enriched expression in *D. melanogaster*. Among these genes, we found *cutoff*, *rhino* and *deadlock* that were recently shown to form a complex initiating the transcription of dual-strand *pi*RNA clusters in *D. melanogaster* (Mohn et al., 2014). All together, these results indicate that enhanced TE activity somehow reinforce sexual selections. First, enhanced TE activity has impacted the female-specific selection on genes associated with the *pi*RNA regulation. Second, TE activity leads to a strong male-specific activity of TEs resulting from a global repression in the ovaries by *pi*RNAs. However, when TE activity is less pronounced, as for instance in *D. simulans*, we found that the female-specific selection on *pi*RNA genes tends to be more relaxed. In the latter case, the relaxation of the selective constraints is correlated with the fact that the transcriptional TE activity is less biased towards the male germline (*i.e.* less *pi*RNA silencing in the ovaries).

Interestingly, the opposite pattern has been observed in mice: the *pi*RNA regulatory pathway is mainly efficient in testes but not in ovaries. The *pi*RNA producing cells are involved in the non-canonical transcription of billions *pi*RNAs and the specific activation of a complex network of pleiotropic genes. This machinery requires large energy costs and thus, was likely selected in only one of the two germlines. The female-biased investment in *Drosophila* may have been favored because of the central importance of maternally inherited *pi*RNAs. The epigenetic transmission of *pi*RNAs and associated regulatory proteins through mothers was shown to be essential for both germline maintenances in the next generation and to ensure to the progeny an optimal fitness (Aravin and Bourc'his, 2008; Molaro et al., 2014; Vagin et al., 2006). In mice, *pi*RNAs derived from TEs have probably been favored during the drastic epigenetic reprogramming, known to be specific to the male germline. This process is also known to be associated with a huge transcriptional activation of *LINE1* retrotransposons (Aravin et al., 2009; Kuramochi-Miyagawa et al., 2008).

Does the TE activity contributes to molecular innovations in testes?

In testes, we have observed a higher TE transcriptional activity for the majority of *D. melanogaster* active TE families compared to ovaries. An overexpression of TEs in the male germline may have important evolutionary consequences participating to the process of “male-biased evolution” of the genomes.

The male-driven evolution has been proposed to describe a general evolutionary phenomenon that underlies mutation rates differences between the male and the female germline. Indeed, it has been frequently

observed in many groups of species that the rate of DNA substitution is more elevated in the male germlines (Connallon and Clark, 2010; Connallon and Knowles, 2005; Hurst and Ellegren, 1998; Parsch and Ellegren, 2013). These differences are generally attributed to the number of germ cell divisions that are higher in males than in females (Wilson Sayres and Makova, 2011). However, the increased TE activity observed, here, in the male germline may constitute an important source of male-biased genetic mutations and rearrangements that can feed the male-biased evolution of species. Interestingly, it has been shown in mammals that the most active TE classes (*LINEs* and *SINEs*) display a male-biased distribution in the genome (i.e. an over-representation on autosomes compared to the X chromosome), associated to testes specific overexpression (Jurka et al., 2002; Nellaker et al., 2012). The specific role of TEs in this context is still poorly investigated, but it would be interesting to test how many species are concerned by a testis-biased activity of TEs and how these TEs are involved in the faster genome evolution driven by the male germline.

In parallel, a growing number of works have noticed that the vast majority of newly emerged genes begin to be expressed in a testes-specific manner (Heinen et al., 2009; Kaessmann, 2010; Levine et al., 2006; Light et al., 2014; Paulding et al., 2003; Ponce and Hartl, 2006; She et al., 2004). This phenomenon, called the “out of testes” hypothesis, uncovers a large spectrum of animal species (Kaessmann, 2010). This hypothesis has been recently extended to plants (“out of pollen” hypothesis, (Wu et al., 2014)). Indeed, most of new genes are TE-derived sequences that start to be expressed in testes (Ruiz-Orera et al., 2015; Wu et al., 2014). In addition, TE transposition are known to play an important role in the generation of new duplicated genes, as well as, new retrotransposed mRNA or non-coding RNAs (Kaessmann, 2010; Wu et al., 2014). TEs are also assumed to provide *cis*-regulatory sequences in order to switch on the transcription of new genes. Indeed, TEs were found to be enriched close to many newly testis-expressed genes (Bennetzen, 2000; Toll-Riera et al., 2009; Wissler et al., 2013).

Here, we have demonstrated that the TE overexpression toward testes is the consequence of a *pi*RNA mediated silencing that is solely efficient in ovaries. Hence, we propose that the male germline might promote the emergence of genetic conflicts through the enhanced activity of TEs. TEs in testes can constitute a force, facilitating new genetic rearrangements, enlarging the selective spectrum required to promote diversifying selection and adaptive innovations.

Conclusion

Here we found first that the tempo and the dynamics of TE invasion are clearly different between two closely related flies: *D. simulans* have experienced ancient waves of TE invasion whereas *D. melanogaster* still undergo recent TE bursts. Second, we proposed that these two different TE invasion histories have strongly affected the host defense machinery involved in TE silencing through the production of divergent pools of *pi*RNAs in ovaries. Third, we found that the “post-invasion” *pi*RNA-mediated response is dramatically enhanced in ovaries and almost absent in testes. From the differences revealed by this species comparison, we proposed a model including both TE and *pi*RNA silencing dynamics that take place through the female germline (Figure 8). When a new emerging TE family invades a population, it starts to transpose with an optimal transposition rate until it get trapped by the host *pi*RNA silencing machinery. In the early stage of the invasion, new TE insertions are characterized by a strong insertion polymorphism, rarely fixed at a given *locus* and as well, rarely targeted by *pi*RNAs. With time, alongside the TE amplification process, the targeted *pi*RNA response becomes active, since some insertions are fixed into *pi*RNA clusters. Highly amplified TEs are thus silenced and copies progressively degenerate. However, while euchromatic copies degenerate, copies embedded into heterochromatic *pi*RNA clusters will persist as fragments and thus maintain *pi*RNA production against the dead TE family.

Materials and methods

Drosophila Stocks

The population samples used for transcriptome and small RNA sequencing analyses were collected from geographically distinct areas. *D. melanogaster* populations were collected from Zimbabwe, Gotheron (France) in addition to the laboratory strain M19 (derived from a *w¹¹¹⁸* strain). *D. simulans* populations were collected from Nairobi (Kenya), Fukuoka (Japan) and Nasrallah (Tunisia). Populations were maintained as isofemale lines at 25°C since the date of capture.

mRNA Library Preparation

Total RNA was extracted from 30 pairs of ovaries and testes, dissected 2 to 4 days after hatching, according to the manufacturer's instructions (Macherey-Nagel). From these samples, PolyA mRNAs were extracted using the "FastTrack MAG Micro mRNA isolation kit" (Life Technologies) according to the recommendations. The resulting mRNA was fragmented with RNA fragmentation reagents (Ambion) according to the kit recommendations. Fragmented RNAs were treated with antarctic phosphatase (NEB) and Polynucleotide Kinase (NEB). Strand orientated libraries were prepared with the "Truseq Small RNA sample prep Kit" (Illumina). The final gel purification step has been replaced by a PCR cleanup with AMPureXP beads (Beckman-Coulter).

Small RNA Library Preparation

Total RNA was extracted from 50 and 100 pairs of ovaries and testes respectively, dissected 2 to 4 days after hatching. A trizol extraction procedure was used according to the manufacturer procedure (TRIzol Reagent, invitrogen). Small RNAs were size fractionated from 1µg total RNA on a TBE-urea 15% acrylamide gel. The resulting RNAs have been treated with the Illumina "Truseq Small RNA sample prep Kit" according to the manufacturer recommendations.

Sequencing

Libraries were sequenced on an Illumina Hiseq 1000 instrument, with a TruSeq SR Cluster Kit v3-cBot-HS (Illumina) and a TruSeq SBS v3-HS - 50 cycles Kit (Illumina), using a Single Read 50 bp recipe. Libraries have been pooled in equimolar proportions and diluted to a final concentration of 12pM, according to Illumina recommendations. The data were demultiplexed using the distribution of CASAVA software (CASAVA-1.8.2) (Mortazavi et al., 2008). The quality of the data was checked with the software FastQC 0.10.1. (Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

Read mapping and Differential Expression Analysis

Single-end reads from each library were filtered with UrQt software (Modolo and Lerat, 2015), only high quality reads were retained (phred score > 33) for the analyses. Illumina adapter was removed using scythe software (<https://github.com/najoshi/>). Only reads with a minimal length of 15 nucleotides were kept. The STAR software (Dobin et al., 2013) was used to map RNA-seq reads on the different databases. TE-derived mRNAs from both *D. melanogaster* and *D. simulans* were identified mapping reads against a custom TE library derived from RepBase (Jurka et al., 2005), that contains consensus sequences of all known TE in *Drosophila* genomes. Only reads mapping to a unique consensus sequence of the RepBase database were kept to generate count table by TE family. Row count matrix were obtained using HTSeq software from the STAR output sam files, and the R bioconductor package DESeq2 (Love et al., 2014) was used for differential expression analyses. The software Deseq2 implements a generalized linear model where counts for each gene in each sample are modeled using a negative binomial distribution. Differentially expressed TEs consensus were selected comparing testes to ovaries conditions in each population (corrected p-value < 0.1 and 10% False Discovery Rate). The base mean values extracted from DESeq2 analyses represent the means of the normalized numbers of reads mapping in TEs between the tested samples (corrected for library size and gene specific coverage). We choose only TEs with a base mean >10, considering that under this value the element is not expressed in gonads.

The same procedure was used for gene-derived mRNAs mapping reads to the *D. melanogaster* (dmel-r6.01 flybase genome) and *D. simulans* (Andolfatto 2012 Genome) reference genomes.

Comparison of genome TE content

Short contigs (<100kb) of the reference genome assemblies of *D. melanogaster* and *D. simulans* (dmelr6.01 and dsim2012) were removed. Comparison between the genome TE content of the two species was carried out using RepeatMasker (<http://www.repeatmasker.org>) with default parameters and RepBase as TE libraries.

Amplification range of TE

From differential expression analyses, subsets of TE consensus sequences were selected and systematically blasted against the reference genomes. BLAST hits with at least 1kb length and 90% of identity to the consensus sequence were kept for the analyses. Pairwise identity of blast hits from each TE consensus family was calculated and classified according to their degree of divergence. Each TE family copy number was plotted according to their level of similarity to the consensus sequence (Supplementary Figure 6). Four qualitative classes were designed according to the maximum copy number of each distribution. TE sequences that count less than 10 (class 10) copies of this kind, between 6 and 20 (class 20), between 21 and 40 (class 40) and more than 40 copies (class >40) that match at different genomic locations. This classification gives a qualitative estimation of the recent TE activity.

Small RNA mapping to RepBase designed class of TE

Barcodes and adapters were removed from both testes and ovaries small RNA libraries using the Cutadapt tool and reads that are between 5 and 45 bp after stripping were kept. TE-derived transcripts were identified by mapping sequenced small RNAs (19 to 30 nt) to a set of selected *Drosophila* TE consensus sequences present in RepBase using bowtie (Langmead et al., 2009) allowing up to 1 mismatch ([– K 1, - 1]) and multiple matches to one position ([–a –best –strata –v 1]). All reads mapping to a unique TE consensus were pooled, and reads mapping to more than one TE consensus were discarded.

The ping-pong signature is the probability that a randomly sampled piRNA from a given TE family have an antisense binding RNA overlapping on the first 10 bp. It was calculated using a custom script as described in (Antoniewski, 2014). For each piRNA (23 to 29 nt), the proportion of overlapping antisense binding partners whose 5' end lies on the first 10 nucleotides was then determined by TE family and by pooled condition. Several sets of small RNA have been tested.

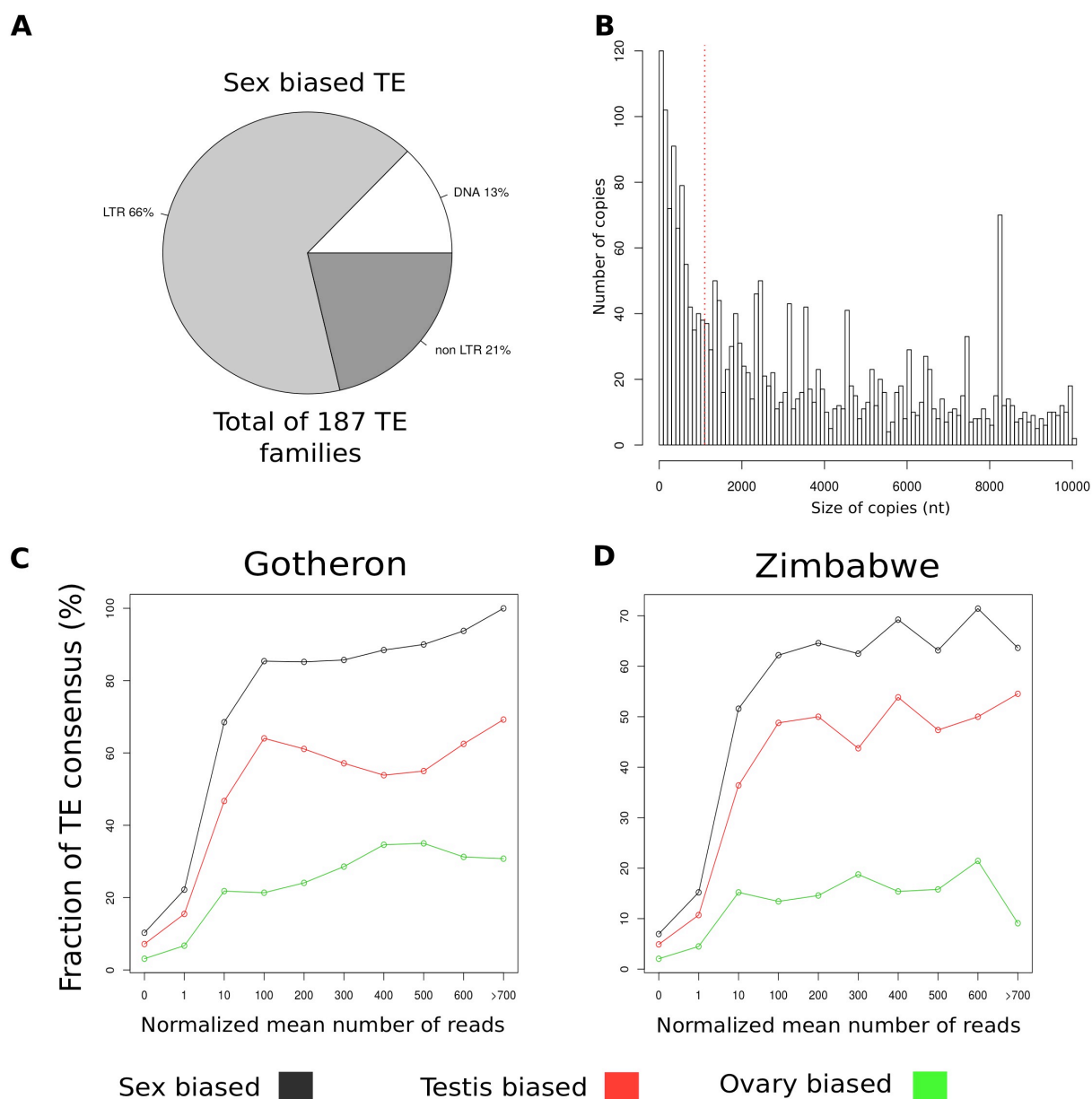


Figure 1. Sex-biased TEs is a genome wide effect in *D. melanogaster*

The pie chart (A) shows the proportion of the three main TE classes among the 187 TE families differentially expressed between male and female germlines of *D. melanogaster* populations ($p\text{-value} < 0.1$, $\text{FDR} = 0.1$). The figure (B) shows the size distribution of TE families that are differentially expressed between the two germlines. The y-axis displays the number of copies found in the *D. melanogaster* reference genome, according to their size in nucleotides (x-axis). Figures (C) and (D) show the proportion of TE consensus displaying an overexpression (y-axis) in males (red) in ovaries (green) and a biased expression in both sexes (black), according to their levels of expression (x-axis). The x-axis displays the range of expression and includes all TE consensus for which the expression levels are greater than or equals to this range. (C) is relative to the Gotheron European population and (D), to the Zimbabwe African population.

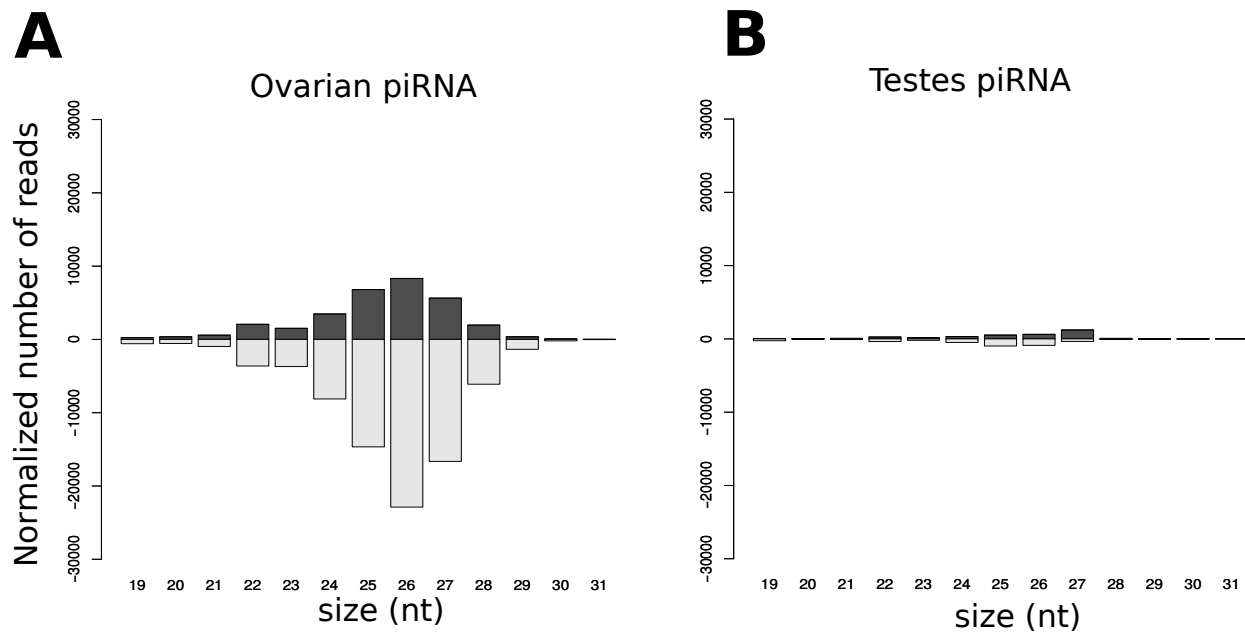


Figure 2. TE silencing is mainly effective in ovaries of *D. melanogaster*

Graphics (A) and (B) display the number of sens RNAs (black) and antisens RNAs (grey) according to their size in nucleotides. Number of reads was normalized by the total number of reads from each library. Figure (A) shows size distributions of small RNAs in ovaries and (B) in testes.

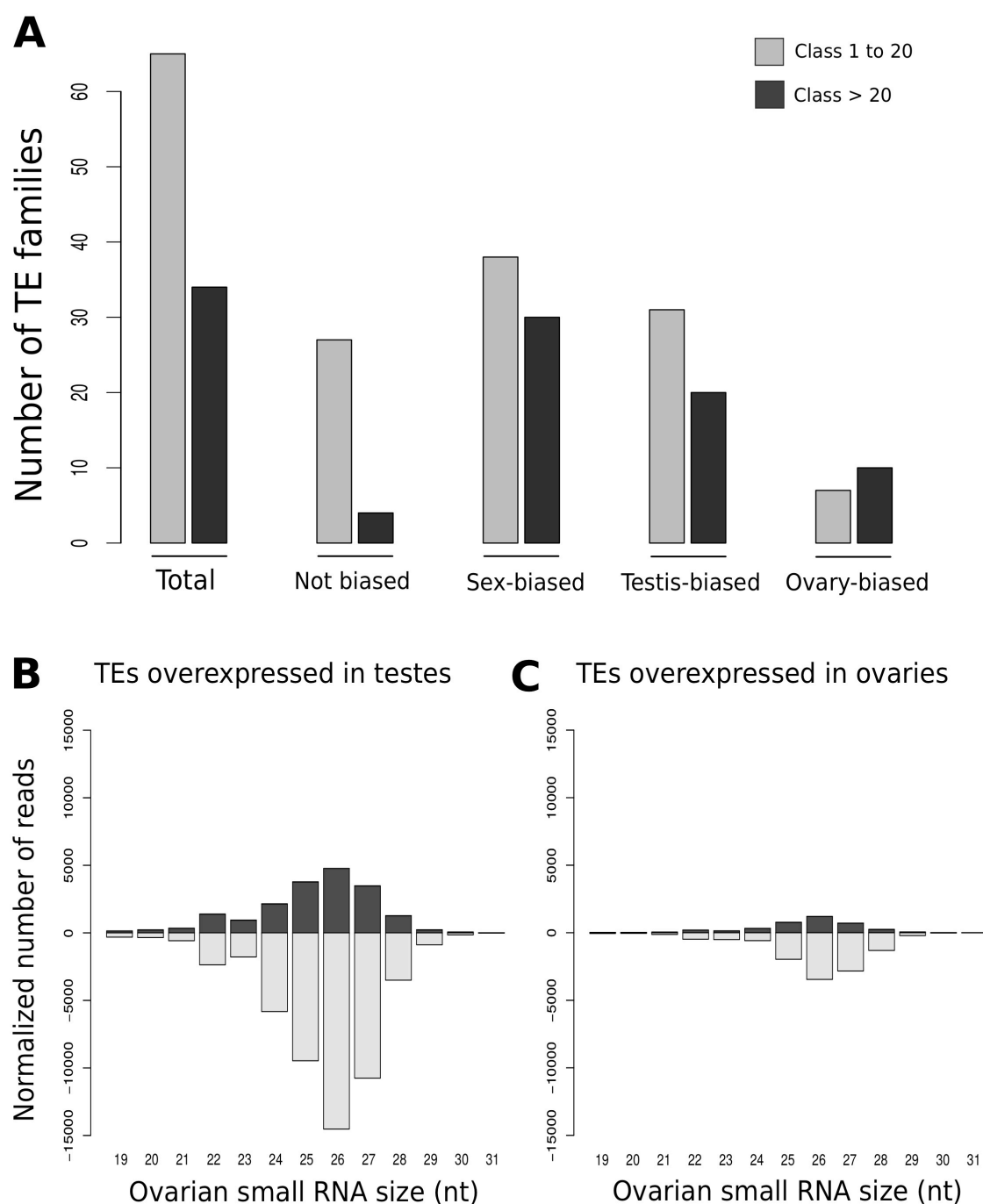


Figure 3. TEs displaying a testis-biased expression are strongly silenced in *D. melanogaster* ovaries

Graphic (A) shows TE family consensus that were classified according to their levels of amplification within *D. melanogaster* reference genome. Grey bars represent TE families that contain between 1 to 20 copies in the genome. TE families found to have more than 20 copies in the reference genome are shown in black. Only TE families expressed in the germline were kept for the analysis (see methods). “Total” represents all the TE families considered: TE families not differentially expressed between the two germlines (not biased), TE families that display a sex-biased expression pattern (Sex-biased) that correspond to TEs overexpressed in testis (Testis-biased) and in ovaries (Ovary-biased). Figure (B) and (C) display the number of sense RNAs (black) and antisense RNAs derived from ovarian libraries (grey) according to their size in nucleotides. Figure (B) displays small RNAs specifically matching with TE families overexpressed in testes and (C) small RNAs specifically matching with TE families overexpressed in ovaries.

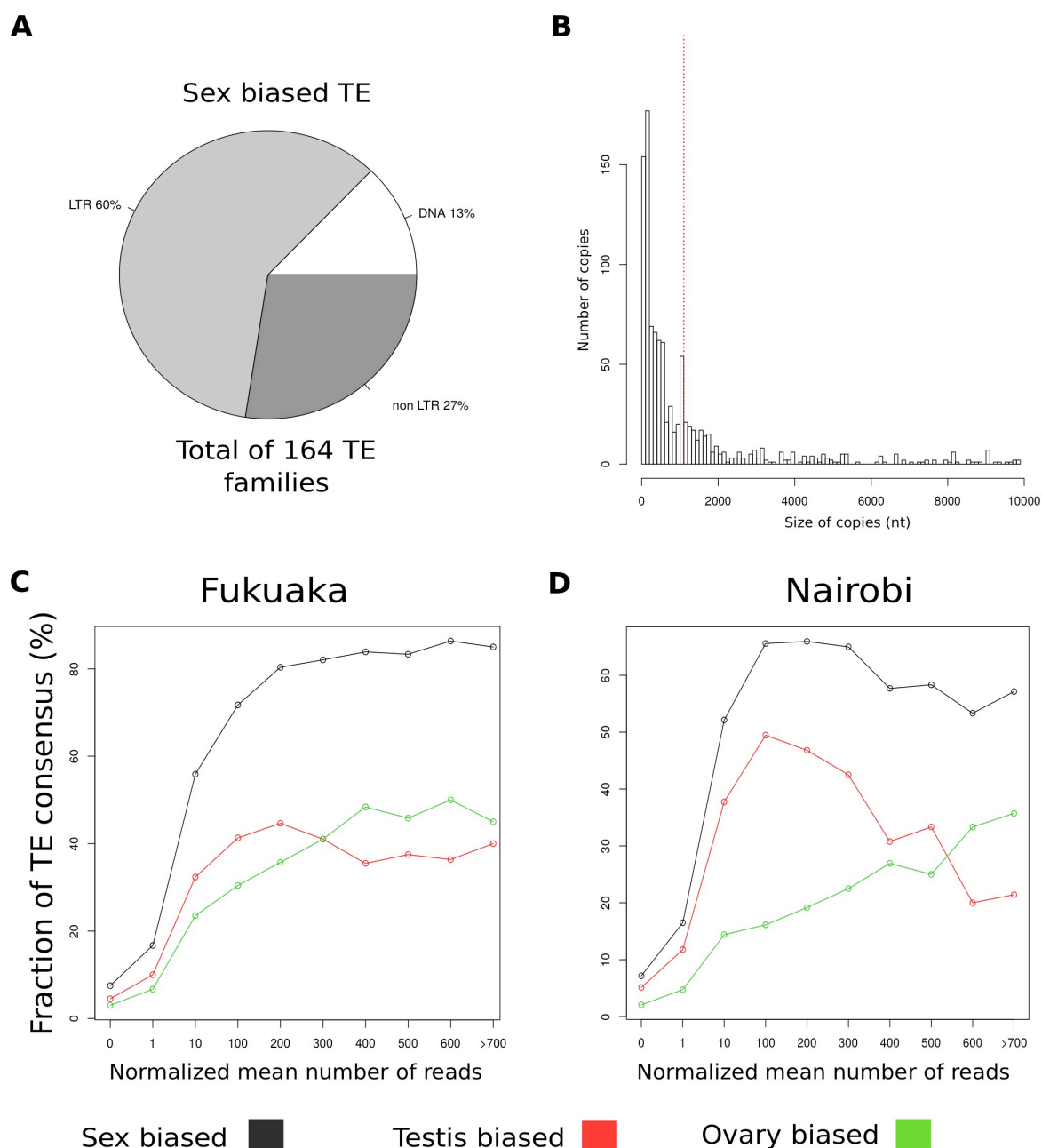


Figure 4. Sex-biased TEs is a genome wide effect in *D. simulans*

The pie chart (A) shows the proportion of the three main TE classes among the 164 TE families differentially expressed between male and female germlines of *D. simulans* populations (p -value <0.1 , FDR = 0.1). The figure (B) shows the size distribution of TE families that are differentially expressed between the two germlines. The y-axis displays the number of copies found in the *D. simulans* reference genome according to their size in nucleotides (x-axis). Figures (C) and (D) show the proportion of TE consensus displaying an overexpression (y-axis) in males (red) in ovaries (green), and a biased-expression in both sexes (black), according to their levels of expression (x-axis). The x-axis displays the range of expression and includes all TE consensus for which the expression levels are greater than or equals to this range. (C) is relative to the Fukuaka Japanese population and (D), to the Nairobi African population.

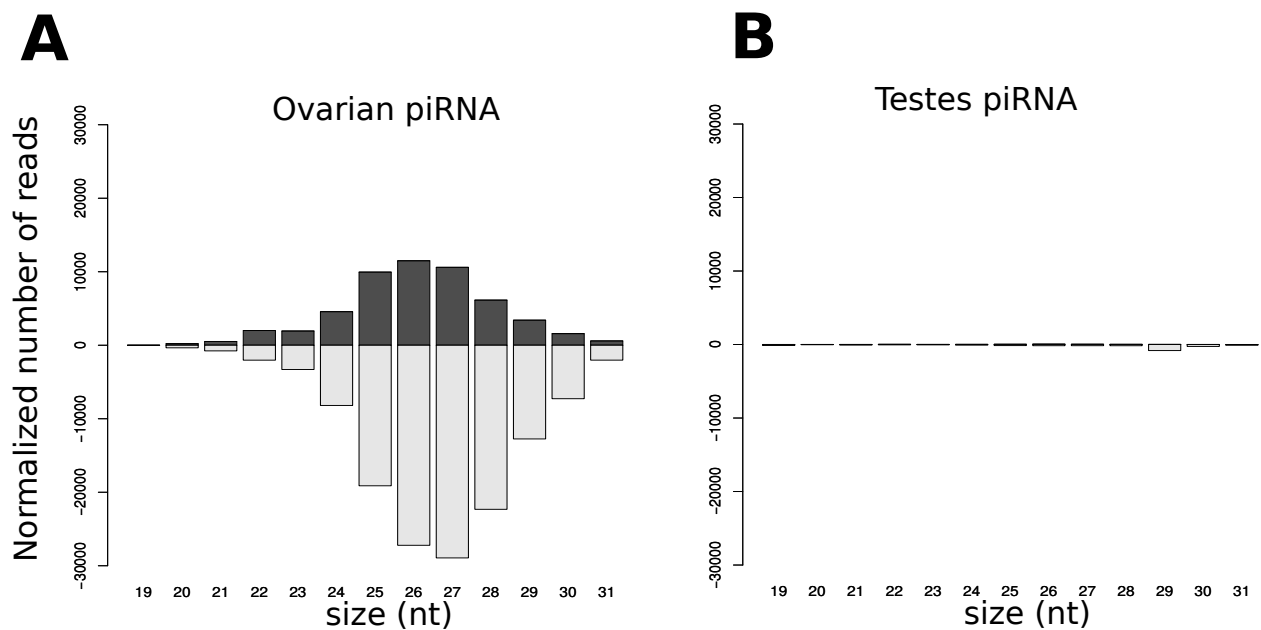


Figure 5. TE silencing is mainly effective in ovaries of *D. simulans*

Graphics (A) and (B) display the number of sens RNAs (black) and antisens RNAs (grey) according to their size in nucleotides. Number of reads were normalized by the total number of reads form each library. (A) shows size distribution of small RNAs in ovaries and (B), in testes.

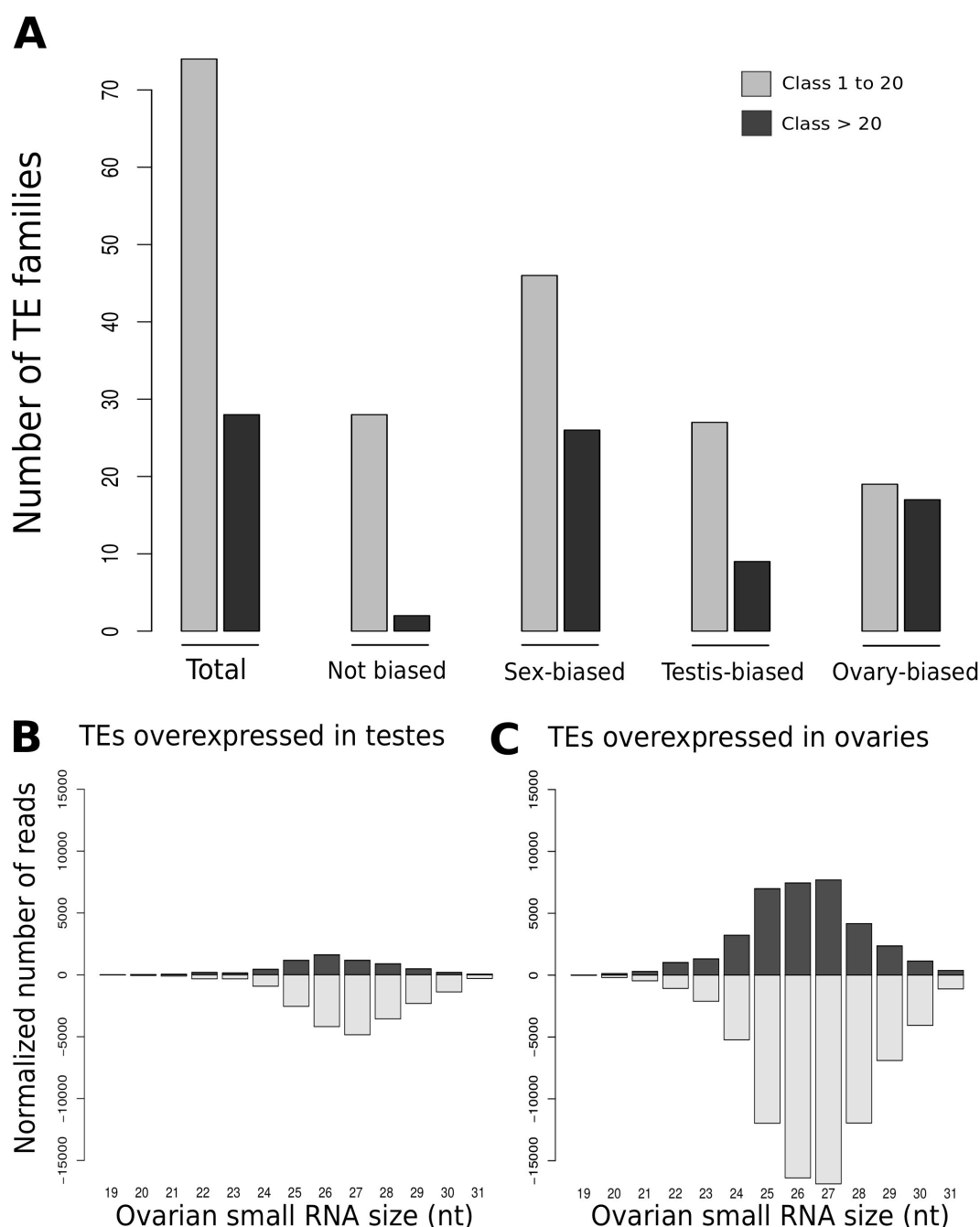


Figure 6. TEs displaying a testis-biased expression are not strongly silenced in *D. simulans* ovaries

Graphic (A) shows TE family's consensus that were classified according to their levels of amplification within *D. simulans* reference genome. Grey bars represent TE families that contain between 1 to 20 copies in the genome. TE families found to have more than 20 copies in the reference genome are shown in black. Only TE families expressed in the germline were kept for the analysis (see methods). "Total" represents all the TE families considered: TE families not differentially expressed between the two germlines ("not biased"), TE families that display a sex biased expression pattern (Sex-biased) that correspond to TE overexpressed in testes ("Testis-biased") and in ovaries ("Ovary-biased"). Figure (B) and (C) display the number of sense RNAs (black) and antisense RNAs derived from ovarian libraries (grey) according to their size in nucleotides. Figure (B) displays small RNAs specifically matching with TE families overexpressed in testes and (C), small RNAs specifically matching with TE families overexpressed in ovaries.

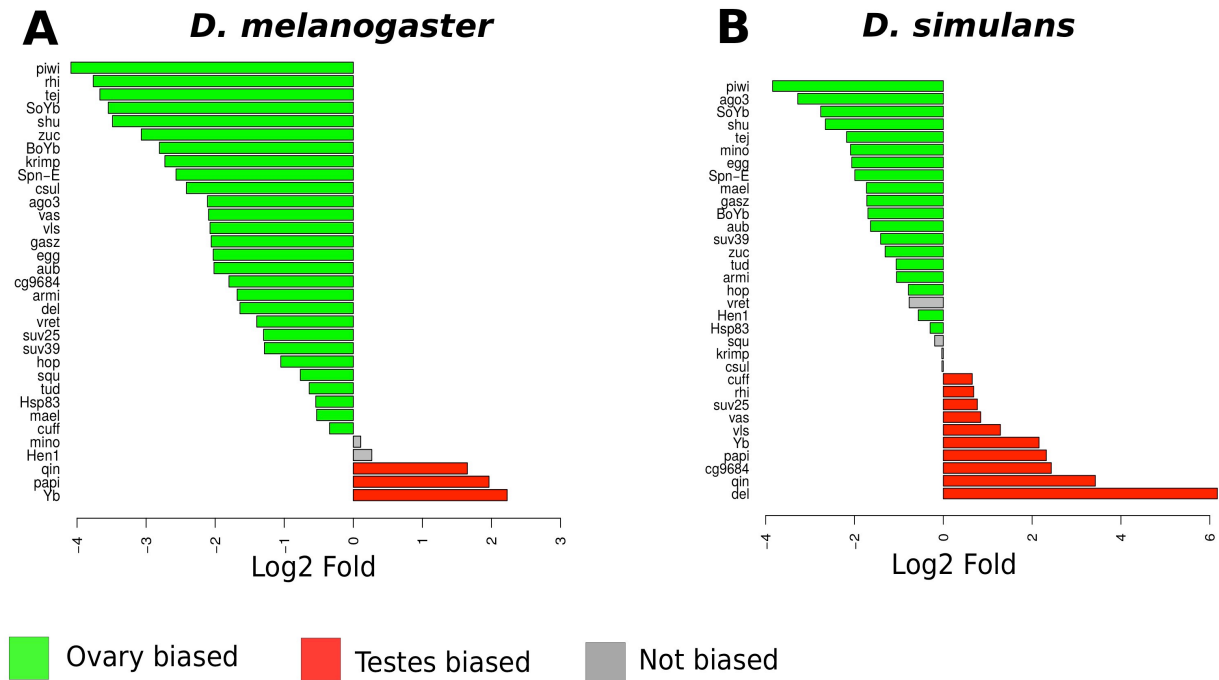


Figure 7. Genes associated to *pi*RNA silencing display female specific selection signatures

The figure shows the list of *D. melanogaster* genes specifically associated to the germline silencing of TEs by the *pi*RNA pathway (A) and the corresponding orthologous genes in *D. simulans* (B). The figure displays the log2 ratio of the testis gene expression on ovary gene expression. Green bars show genes overexpressed in ovaries. Red bars show genes overexpressed in testes. Genes that are not significantly differentially expressed are shown in grey.

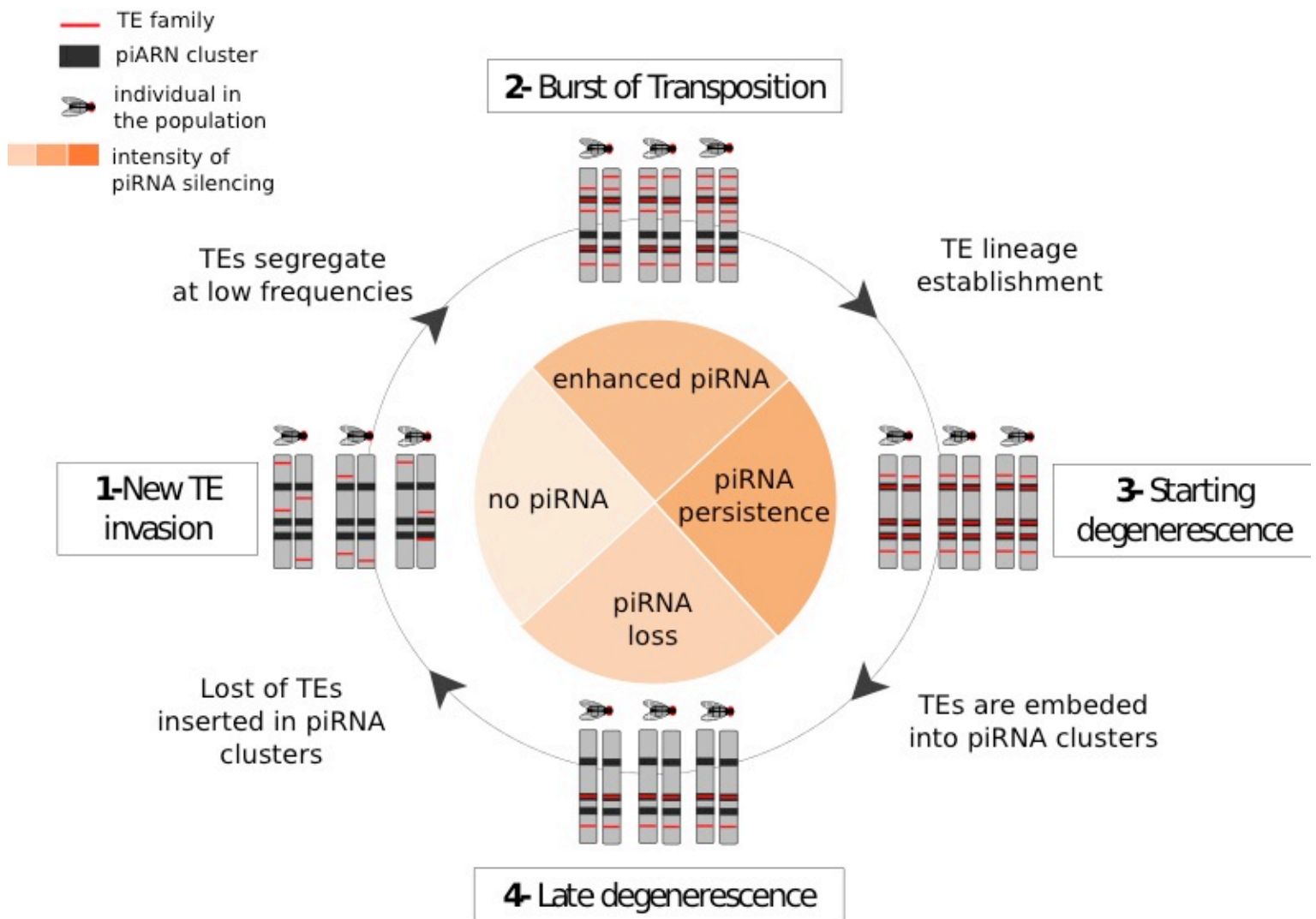
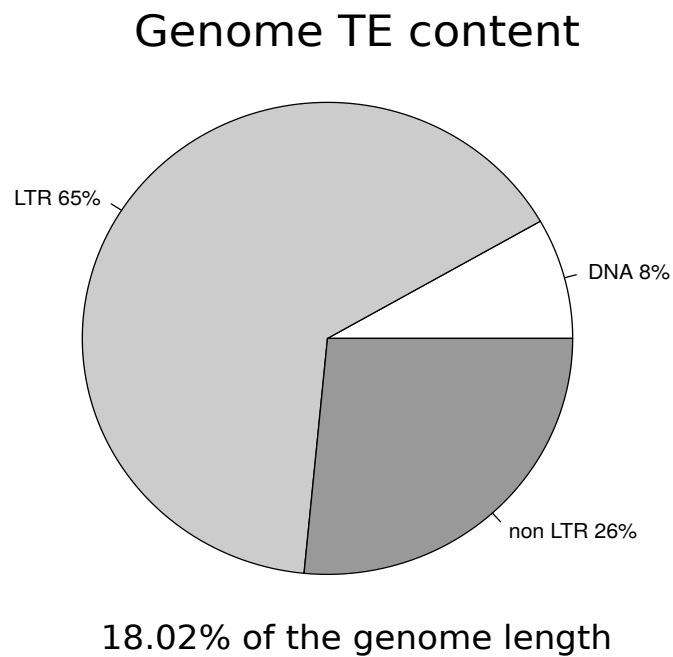


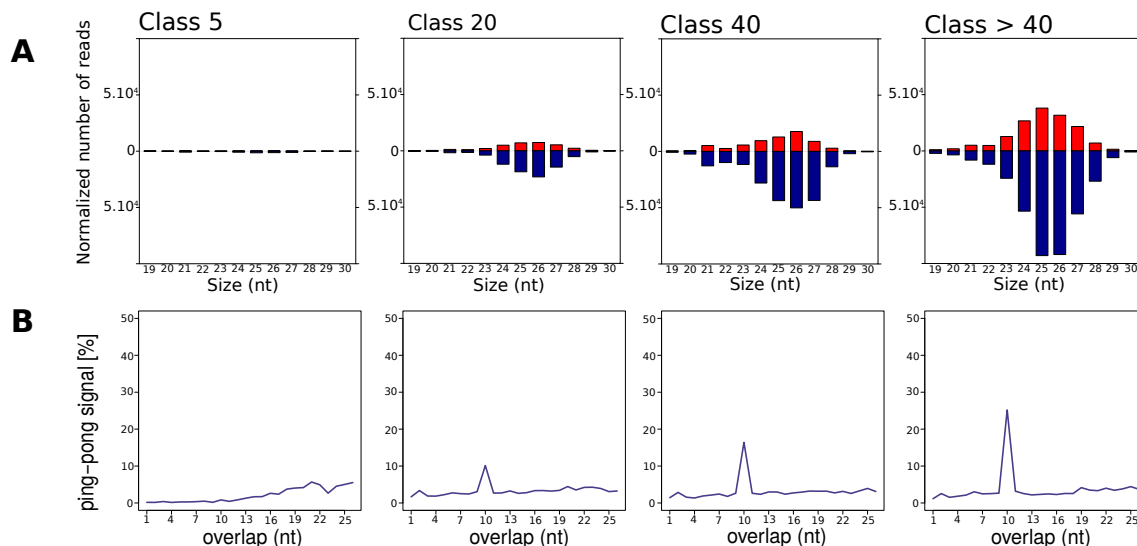
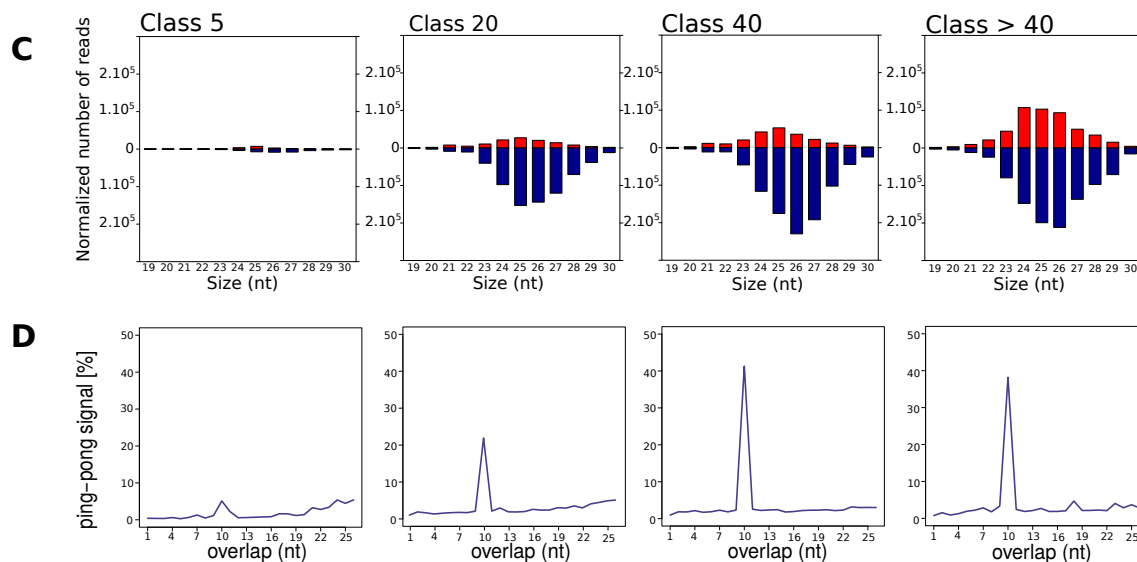
Figure 8. Evolutionary dynamics of a new TE emerging family under control of the *piRNA* silencing pathway

The step 1 represents a young emerging TE (red line) emerging in a population of diploid genomes (pair of grey bars). The activity of the new element is associated to a strong insertion polymorphism and thus, present at diverse genomic locations that are not fixed yet in the population. The step 2 corresponds to the establishment of the TE family within the population of genomes. Following the TE burst of amplification, some TE insertions are now found to be fixed in many genomic locations. However, in response to the strong amplification wave, some TEs inserted into *piRNAs* producing regions (black line) have been selected, in order to limit the expansion of the highly amplified TE family. The step 3 corresponds to the long-term establishment of the TE family, where most of the TE insertions are now fixed and not able to transpose anymore because of multiple *piRNA* producing copies inserted into *piRNA* clusters. The *piRNA* response will persist until the step 4. The step 4 corresponds to the very long-term dynamics where most of the TE insertions are found completely degenerated and fragmented. TEs are progressively removed from the genome, and the loss of ancient fragmented copies inserted in *piRNA* cluster goes along with the progressive loss of the genomic *piRNA* response. Once completely lost, the cycle is over and a new reinvasion can occur.



Supplementary Figure 1

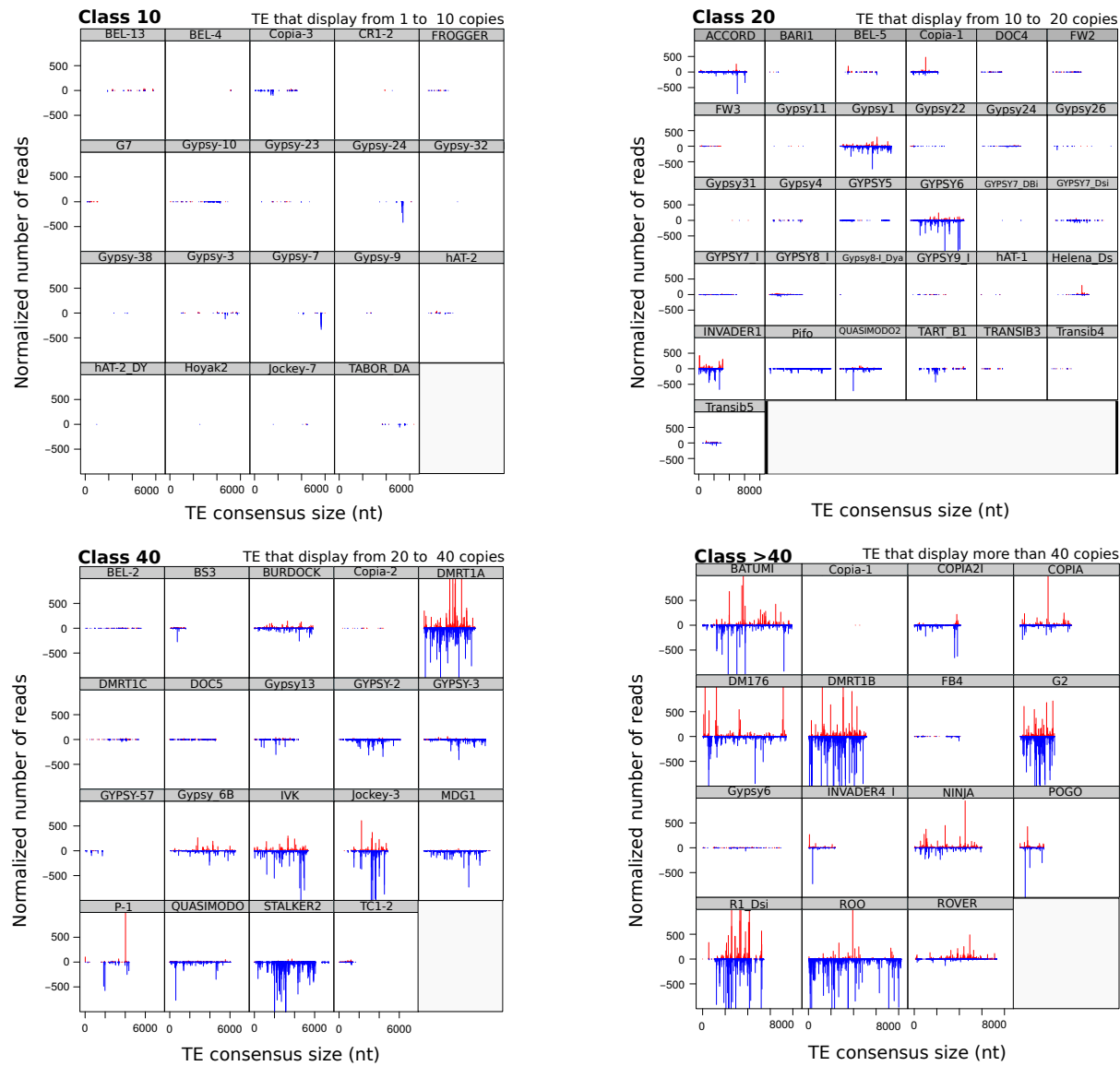
The pie chart (A) shows the proportion of the different types of TEs in the *D. melanogaster* genome. TEs represent 18 % of *D. melanogaster* genome length. DNA transposons (DNA) are shown in white, non-LTR retrotransposons (non LTR) are shown in dark grey, and LTR retrotransposons (LTR) are shown in grey.

D. melanogaster*D. simulans***Supplementary Figure 2**

The figure (A) shows the number of reads matching with TE consensus families according to their size in nucleotides in *D. melanogaster*. TE families were classified according to their level of amplification within the reference genome. “Class 10” represents TE families that are found between 1 and 10 copies, “Class 20” for TEs found between 11 and 20 copies, “Class 40” between 21 to 40 copies and “Class >40” TE families found in more than 40 copies. The figure (B) shows the proportion of reads that overlap (y-axis) according to the starting position of the overlap on 3’ reads (x-axis) in *D. melanogaster*. The ping-pong signature (B) was displayed according TE families class of amplification.

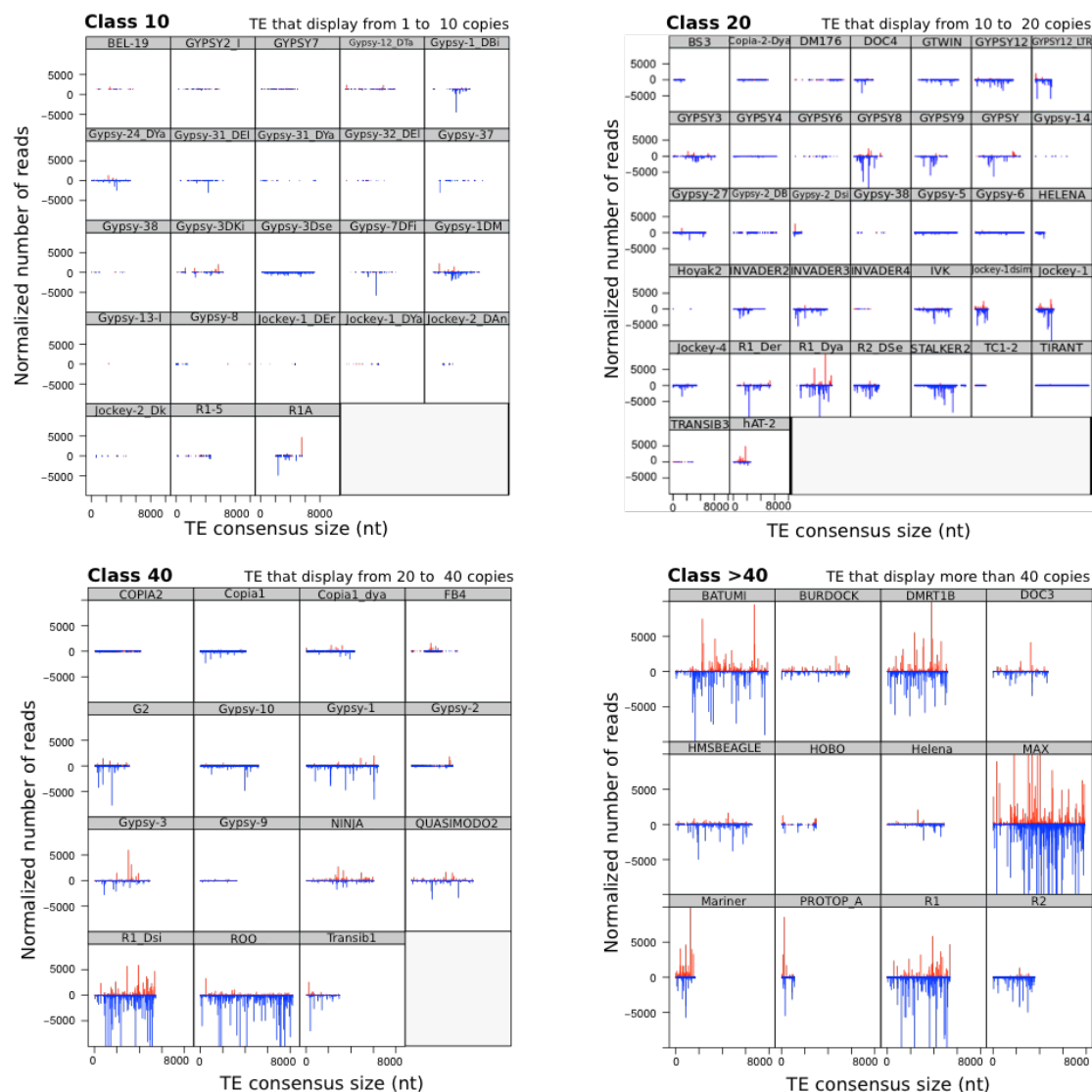
The figure (C) shows the number of reads matching with TE consensus families according to their size in nucleotides in *D. simulans*. TE families were classified according to their level of amplification within the reference genome. The figure (D) shows the proportion of reads that overlap (y-axis) according to the starting position of the overlap on 3’ reads (x-axis) in *D. simulans*. The ping-pong signature (D) was displayed according TE families class of amplification.

D. melanogaster



Supplementary Figure 3a

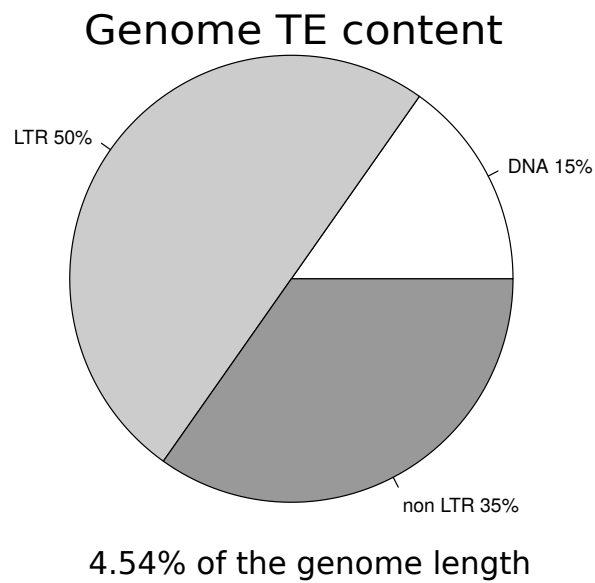
D. simulans



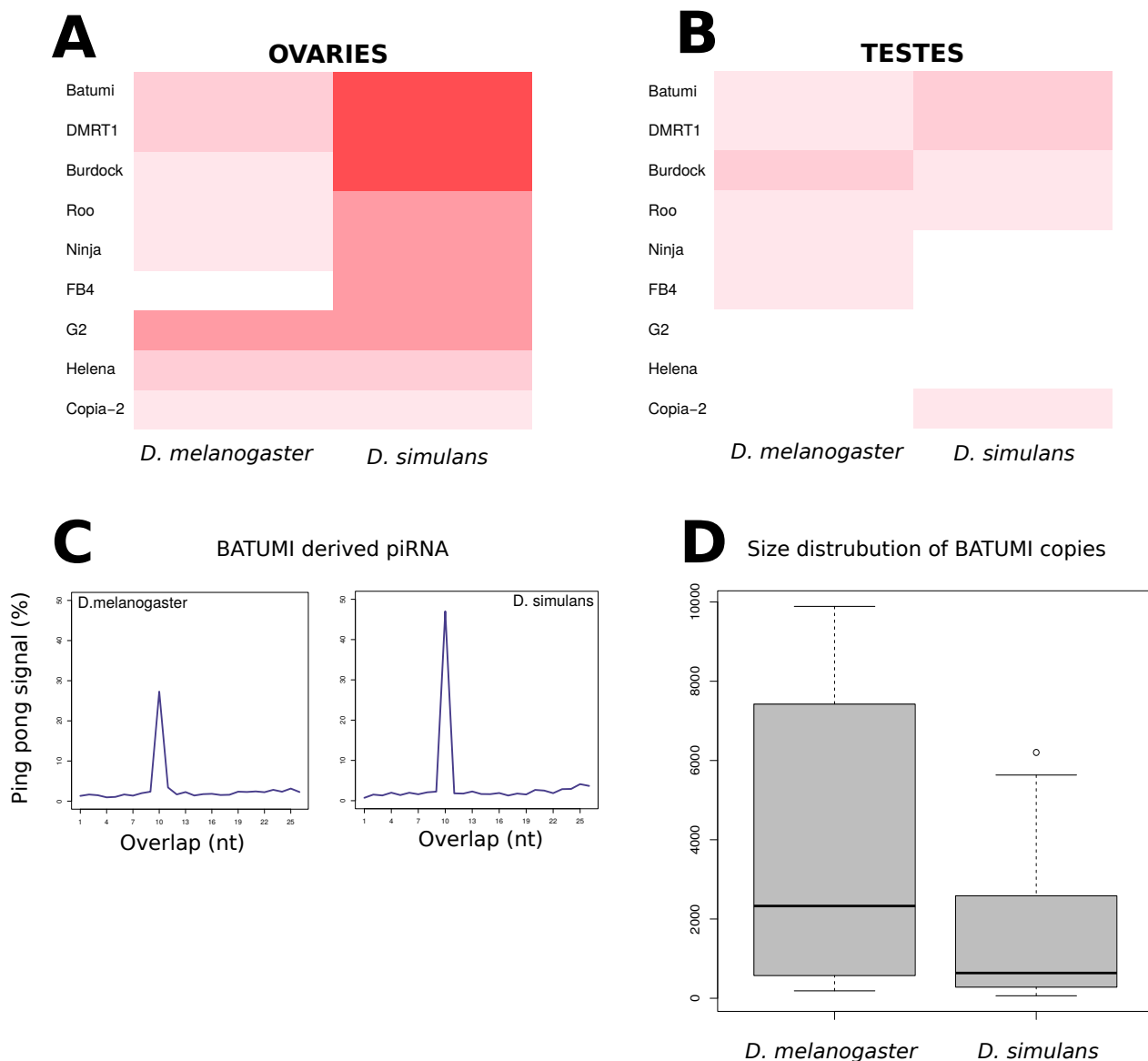
Supplementary Figure 3b

Graphics on Supplementary Figure 3a, display the number of *pi*RNA (y-axis) mapping by nucleotide position along the TE consensus sequence (x-axis) in *D. melanogaster*. The presented TE families were classified in four tables that correspond to their level of amplification (Class 10, 20, 40, >40) as described in Supplementary Figure 2.

Graphics on Supplementary Figure 3b, display the number of *pi*RNA (y-axis) mapping by nucleotide position along the TE consensus sequence (x axis), in *D. simulans*. The presented TE families were classified in four tables that correspond to their level of amplification (Class 10, 20, 40, >40) as described in Supplementary Figure 2.

**Supplementary Figure 4**

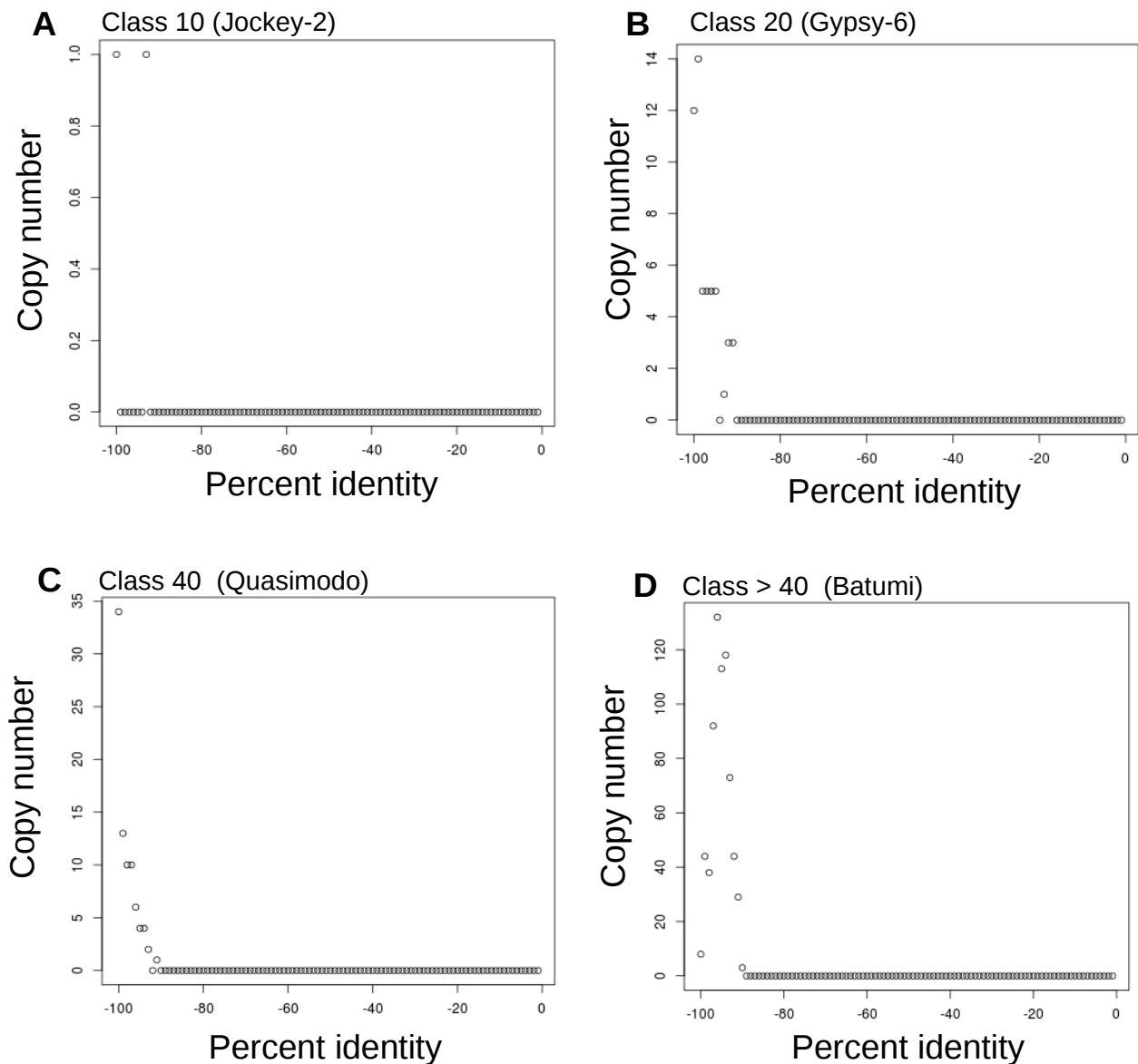
The pie chart shows the proportion of the different types of TEs in the *D. simulans* genome. TEs represent 4.5 % of *D. simulans* genome length. DNA transposons (DNA) are shown in white, non-LTR retrotransposons (non LTR) are in dark grey, and LTR retrotransposons (LTR) are shown in grey.



Supplementary Figure 5

Supplementary Figure 5A and B, show the ping-pong signatures for a subset of TE families sharing the same levels of amplification in both *D. melanogaster* and *D. simulans* species. The heatmap shows in (A) displays the ping-pong signature for each TE families from ovarian *piARN* libraries. The heatmap presented in (B) shows the ping-pong signature for each TE families, from testes *piARN* libraries.

The intensity of the ping-pong signal is represented as color gradient from white to red (from white when signal is absent to red when more than 50 % of the reads present the overlap signature at position 10 of the read). Figure 5C, specifically displays the percentage of overlapping reads for *piRNA* derived from BATUMI LTR retrotransposon. The y-axis displays proportion of overlap according to their starting position on 3' reads (x-axis). Figure 5D, shows the size distribution in nucleotides of the BATUMI copies (y-axis) found in *D. melanogaster* (left) and *D. simulans* (right).



Supplementary Figure 6

The figure shows the plot distribution of TE copy number according to the percentage of identity to the consensus sequence. The maximum of each distribution was used to classify TE according to their amplification range. Panel (A) shows the “class 10” TEs that amplified up to 10 copies. Panel (B) is the “class 20” for TEs that amplified up to 20 copies. Panel (C) shows the class 40 (up to 40 copies) and panel (D) shown the class >40 (more than 40 copies).

Bibliographie

- Akkouche, A., T. Grentzinger, M. Fablet, C. Armenise, N. Burlet, V. Braman, S. Chambeyron, and C. Vieira. 2013. Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. *EMBO Rep.* 14:458-464.
- Antoniewski, C. 2014. Computing siRNA and piRNA overlap signatures. *Methods Mol Biol.* 1173:135-146.
- Aravin, A., and T. Tuschl. 2005. Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett.* 579:5830-5840.
- Aravin, A.A., and D. Bourc'h. 2008. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev.* 22:970-975.
- Aravin, A.A., G.J. Hannon, and J. Brennecke. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science.* 318:761-764.
- Aravin, A.A., G.W. van der Heijden, J. Castaneda, V.V. Vagin, G.J. Hannon, and A. Bortvin. 2009. Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet.* 5:e1000764.
- Bartolome, C., X. Bello, and X. Maside. 2009. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol.* 10:R22.
- Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol.* 42:251-269.
- Bergman, C.M., and D. Bensasson. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 104:11340-11345.
- Berry, B., S. Deddouche, D. Kirschner, J.L. Imler, and C. Antoniewski. 2009. Viral suppressors of RNA silencing hinder exogenous and endogenous small RNA pathways in *Drosophila*. *PLoS One.* 4:e5866.
- Blumenstiel, J.P. 2011. Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet.* 27:23-31.
- Bourque, G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev.* 19:607-612.
- Bowen, N.J., and J.F. McDonald. 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* 11:1527-1540.
- Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell.* 128:1089-1103.
- Brennecke, J., C.D. Malone, A.A. Aravin, R. Sachidanandam, A. Stark, and G.J. Hannon. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science.* 322:1387-1392.
- Bucheton, A., J.M. Lavigne, G. Picard, and P. L'Heritier. 1976. Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variations in the efficiency of inducer and reactive strains. *Heredity (Edinb).* 36:305-314.
- Castaneda, J., P. Genzor, and A. Bortvin. 2011. piRNAs, transposon silencing, and germline genome integrity. *Mutat Res.* 714:95-104.
- Chambeyron, S., A. Popkova, G. Payen-Groschene, C. Brun, D. Laouini, A. Pelisson, and A. Bucheton. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci U S A.* 105:14964-14969.
- Connallon, T., and A.G. Clark. 2010. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution.* 64:3417-3442.
- Connallon, T., and L.L. Knowles. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet.* 21:495-499.

- Craig, N.L.C., R.; Gellert, M.; Lambowitz, A. M. 2002. Mobile DNA II. *ASM Press*.
- Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15-21.
- Dolgin, E.S., and B. Charlesworth. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*. 178:2169-2177.
- Ellegren, H., and J. Parsch. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet*. 8:689-698.
- Erwin, A.A., M.A. Galdos, M.L. Wickersheim, C.C. Harrison, K.D. Marr, J.M. Colicchio, and J.P. Blumenstiel. 2015. piRNAs Are Associated with Diverse Transgenerational Effects on Gene and Transposon Expression in a Hybrid Dysgenic Syndrome of *D. virilis*. *PLoS Genet*. 11:e1005332.
- Fablet, M., A. Akkouche, V. Braman, and C. Vieira. 2014. Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis. *Gene*. 537:149-153.
- Gilbert, C., S. Schaack, J.K. Pace, 2nd, P.J. Brindley, and C. Feschotte. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature*. 464:1347-1350.
- Gonzalez, J., K. Lenkov, M. Lipatov, J.M. Macpherson, and D.A. Petrov. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol*. 6:e251.
- Grentzinger, T., C. Armenise, C. Brun, B. Mugat, V. Serrano, A. Pelisson, and S. Chambeyron. 2012. piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res*. 22:1877-1888.
- Gunawardane, L.S., K. Saito, K.M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M.C. Siomi. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*. 315:1587-1590.
- Handler, D., K. Meixner, M. Pizka, K. Lauss, C. Schmied, F.S. Gruber, and J. Brennecke. 2013. The genetic makeup of the *Drosophila* piRNA pathway. *Mol Cell*. 50:762-777.
- Heger, A., and C.P. Ponting. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res*. 17:1837-1849.
- Heinen, T.J., F. Staubach, D. Haming, and D. Tautz. 2009. Emergence of a new gene from an intergenic region. *Curr Biol*. 19:1527-1531.
- Houwing, S., L.M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D.V. Filippov, H. Blaser, E. Raz, C.B. Moens, R.H. Plasterk, G.J. Hannon, B.W. Draper, and R.F. Ketting. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*. 129:69-82.
- Hurst, L.D., and H. Ellegren. 1998. Sex biases in the mutation rate. *Trends Genet*. 14:446-452.
- Jurka, J., V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, and J. Walichiewicz. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 110:462-467.
- Jurka, J., M. Krnjajic, V.V. Kapitonov, J.E. Stenger, and O. Kokhanyy. 2002. Active Alu elements are passed primarily through paternal germlines. *Theor Popul Biol*. 61:519-530.
- Kaessmann, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 20:1313-1326.
- Kawaoka, S., N. Hayashi, Y. Suzuki, H. Abe, S. Sugano, Y. Tomari, T. Shimada, and S. Katsuma. 2009. The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA*. 15:1258-1264.
- Kelleher, E.S., and D.A. Barbash. 2013. Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense. *Mol Biol Evol*. 30:1816-1829.
- Kelleher, E.S., N.B. Edelman, and D.A. Barbash. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol*. 10:e1001428.
- Khurana, J.S., J. Wang, J. Xu, B.S. Koppetsch, T.C. Thomson, A. Nowosielska, C. Li, P.D. Zamore, Z. Weng, and W.E. Theurkauf. 2011. Adaptation to P element transposon invasion

- in *Drosophila melanogaster*. *Cell*. 147:1551-1563.
- Kidwell, M.G., J.F. Kidwell, and J.A. Sved. 1977. Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics*. 86:813-833.
- Kidwell, M.G., and D.R. Lisch. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*. 55:1-24.
- Kofler, R., V. Nolte, and C. Schlotterer. 2015. Tempo and Mode of Transposable Element Activity in *Drosophila*. *PLoS Genet*. 11:e1005406.
- Kolaczowski, B., D.N. Hupalo, and A.D. Kern. 2011. Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol Biol Evol*. 28:1033-1042.
- Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T.W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, and T. Nakano. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev*. 22:908-917.
- Langley, C.H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genet Res*. 52:223-235.
- Langley, C.H., K. Stevens, C. Cardeno, Y.C. Lee, D.R. Schrider, J.E. Pool, S.A. Langley, C. Suarez, R.B. Corbett-Detig, B. Kolaczowski, S. Fang, P.M. Nista, A.K. Holloway, A.D. Kern, C.N. Dewey, Y.S. Song, M.W. Hahn, and D.J. Begun. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*. 192:533-598.
- Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10:R25.
- Lau, N.C., A.G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D.P. Bartel, and R.E. Kingston. 2006. Characterization of the piRNA complex from rat testes. *Science*. 313:363-367.
- Le Rouzic, A., and G. Deceliere. 2005. Models of the population genetics of transposable elements. *Genet Res*. 85:171-181.
- Lee, Y.C., and C.H. Langley. 2012. Long-term and short-term evolutionary impacts of transposable elements on *Drosophila*. *Genetics*. 192:1411-1432.
- Lerat, E., N. Burlet, C. Biemont, and C. Vieira. 2011. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene*. 473:100-109.
- Levine, M.T., C.D. Jones, A.D. Kern, H.A. Lindfors, and D.J. Begun. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A*. 103:9935-9939.
- Light, S., W. Basile, and A. Elofsson. 2014. Orphans and new gene origination, a structural and evolutionary perspective. *Curr Opin Struct Biol*. 26:73-83.
- Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.
- Lu, J., and A.G. Clark. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res*. 20:212-227.
- Malone, C.D., and G.J. Hannon. 2009. Small RNAs as guardians of the genome. *Cell*. 136:656-668.
- Modolo, L., and E. Lerat. 2015. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. *BMC Bioinformatics*. 16:137.
- Mohn, F., G. Sienski, D. Handler, and J. Brennecke. 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell*. 157:1364-1379.
- Molaro, A., I. Falciani, E. Hodges, A.A. Aravin, K. Marran, S. Rafii, W.R. McCombie, A.D. Smith, and G.J. Hannon. 2014. Two waves of de novo methylation during mouse germ cell development. *Genes Dev*. 28:1544-1549.
- Montgomery, E., B. Charlesworth, and C.H. Langley. 1987. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila*

- melanogaster. *Genet Res.* 49:31-41.
- Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 5:621-628.
- Nagao, A., T. Mituyama, H. Huang, D. Chen, M.C. Siomi, and H. Siomi. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. *RNA.* 16:2503-2515.
- Nellaker, C., T.M. Keane, B. Yalcin, K. Wong, A. Agam, T.G. Belgard, J. Flint, D.J. Adams, W.N. Frankel, and C.P. Ponting. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 13:R45.
- Nishida, K.M., K. Saito, T. Mori, Y. Kawamura, T. Nagami-Okada, S. Inagaki, H. Siomi, and M.C. Siomi. 2007. Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA.* 13:1911-1922.
- Obbard, D.J., K.H. Gordon, A.H. Buck, and F.M. Jiggins. 2009a. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci.* 364:99-115.
- Obbard, D.J., J.J. Welch, K.W. Kim, and F.M. Jiggins. 2009b. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5:e1000698.
- Pane, A., K. Wehr, and T. Schupbach. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell.* 12:851-862.
- Parsch, J., and H. Ellegren. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14:83-87.
- Paulding, C.A., M. Ruvolo, and D.A. Haber. 2003. The Tre2 (USP6) oncogene is a hominoid-specific gene. *Proc Natl Acad Sci U S A.* 100:2507-2511.
- Petrov, D.A., Y.T. Aminetzach, J.C. Davis, D. Bensasson, and A.E. Hirsh. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol.* 20:880-892.
- Ponce, R., and D.L. Hartl. 2006. The evolution of the novel Sdic gene cluster in *Drosophila melanogaster*. *Gene.* 376:174-183.
- Robine, N., N.C. Lau, S. Balla, Z. Jin, K. Okamura, S. Kuramochi-Miyagawa, M.D. Blower, and E.C. Lai. 2009. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol.* 19:2066-2076.
- Ruiz-Orera, J., J. Hernandez-Rodriguez, C. Chiva, E. Sabido, I. Kondova, R. Bontrop, T. Marques-Bonet, and M.M. Alba. 2015. Origins of De Novo Genes in Human and Chimpanzee. *PLoS Genet.* 11:e1005721.
- Sanchez-Gracia, A., X. Maside, and B. Charlesworth. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends Genet.* 21:200-203.
- She, X., J.E. Horvath, Z. Jiang, G. Liu, T.S. Furey, L. Christ, R. Clark, T. Graves, C.L. Gulden, C. Alkan, J.A. Bailey, C. Sahinalp, M. Rocchi, D. Haussler, R.K. Wilson, W. Miller, S. Schwartz, and E.E. Eichler. 2004. The structure and evolution of centromeric transition regions within the human genome. *Nature.* 430:857-864.
- Siomi, M.C., K. Saito, and H. Siomi. 2008. How selfish retrotransposons are silenced in *Drosophila* germline and somatic cells. *FEBS Lett.* 582:2473-2478.
- Song, J., J. Liu, S.L. Schnakenberg, H. Ha, J. Xing, and K.C. Chen. 2014. Variation in piRNA and transposable element content in strains of *Drosophila melanogaster*. *Genome Biol Evol.* 6:2786-2798.
- Szakmary, A., M. Reedy, H. Qi, and H. Lin. 2009. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster*. *J Cell Biol.* 185:613-627.
- Toll-Riera, M., N. Bosch, N. Bellora, R. Castelo, L. Armengol, X. Estivill, and M.M. Alba. 2009. Origin of primate orphan genes: a comparative genomics approach. *Mol Biol Evol.* 26:603-612.
- Vagin, V.V., A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P.D. Zamore. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science.* 313:320-324.

- Vieira, C., M. Fablet, E. Lerat, M. Boulesteix, R. Rebollo, N. Burlet, A. Akkouche, B. Hubert, H. Mortada, and C. Biemont. 2012. A comparative analysis of the amounts and dynamics of transposable elements in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. *J Environ Radioact.* 113:83-86.
- Vieira, C., D. Lepetit, S. Dumont, and C. Biemont. 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol.* 16:1251-1255.
- Wilson Sayres, M.A., and K.D. Makova. 2011. Genome analyses substantiate male mutation bias in many species. *Bioessays.* 33:938-945.
- Wissler, L., J. Gadau, D.F. Simola, M. Helmkampf, and E. Bornberg-Bauer. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol Evol.* 5:439-455.
- Wu, D.D., X. Wang, Y. Li, L. Zeng, D.M. Irwin, and Y.P. Zhang. 2014. "Out of pollen" hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biol Evol.* 6:2822-2829.

Discussions

I- Modèle des dynamiques d'ET et de la réponse génomique associée

1- Un équilibre dynamique entre allèles répresseurs et copies actives

Nous avons montré que l'élément *mos1* avait la capacité de transposer lorsqu'il était introduit dans une nouvelle espèce (la lignée M19 de l'espèce *D. melanogaster*). L'élément s'est amplifié à hauteur de 20 copies dans le génome et, depuis plusieurs années, le nombre de copies est demeuré stable. Par ailleurs, nous avons montré que désormais, l'élément *mos1* était sous le contrôle de la voie des *piARN*, suggérant l'implication de cette voie pour limiter l'expansion de *mos1*, ainsi que son maintien à l'état d'équilibre dans le génome. L'élément *mos1* est, d'une part, sensible aux mutations des gènes impliqués dans cette voie, d'autre part, on observe la production *de novo* de *piARN* double brin dérivant de l'élément. Dans la lignée germinale, la régulation de l'élément *mos1* semble être principalement sous contrôle de la voie des *piARN* puisque que les mutants associés à la voie des *siARN* n'affectent pas son expression et on n'observe aucun signe de production de *siARN* dérivés de l'élément.

Par ailleurs, nous avons montré que la dynamique de *mos1* dans le génome semble être affectée par la production de *piARN* principalement dans l'ovaire. En effet, la lignée M19 a été croisée avec une souche vide, successivement sur 6 générations en apportant systématiquement les *piARN* par la mère. On observe *in fine*, la perte quasi totale de l'élément, suggérant que la capacité de l'élément à transposer est fortement diminuée. Au contraire, dans le croisement réciproque, lorsque les *piARN* ne sont pas transmis, l'élément est toujours capable de transposer. Ces résultats suggèrent donc que le *pool* de *piARN* dérivé de *mos1* transmis par la mère à l'ovocyte est nécessaire à l'établissement d'une répression efficace de l'élément. De même, le maintien à l'état d'équilibre du nombre de copies dans le génome, semble assuré par la production de *piARN* maternels qui initient la répression de *mos1* pour la génération suivante. Parmi les 20 copies présentes dans la lignée M19, nous avons identifié différents types d'insertions. Une partie des insertions semble fixée dans le génome tandis que l'autre partie ségrège à plus faible fréquence dans la lignée, témoignant d'une activité récente de l'élément. Parmi ces copies, certaines sont insérées dans des régions riches en ET au niveau desquelles un grand nombre de *piARN* sont produits, tandis que d'autres sont insérées à proximité de régions codantes et ne présentent aucune trace de production de *piARN*.

Ces résultats montrent que l'élément *mos1* est toujours actif dans la lignée M19. Néanmoins, l'activité de l'élément est clairement contrôlée par la voie des *piARN* et probablement initiée à partir de plusieurs insertions de *mos1* dans des régions productrices de *piARN*. La distribution des copies dans le génome est caractérisée par un polymorphisme d'insertion assez élevé, malgré son étonnante stabilité en termes de nombre de copies. Cela

renforce l'idée que l'élément transpose toujours mais que sa stabilisation dans le génome est assurée par un pool de copies productrices de *piARN* qui tamponnent la mobilisation des autres copies dans le génome.

2- Le polymorphisme d'expression des gènes de la voie des *piARN* affecte l'activité de *mos1* dans les populations colonisatrices

L'élément *mos1* est caractérisé par d'importantes variations d'activité entre populations naturelles de l'espèce *D. simulans*. Ainsi, l'élément présente une activité plus importante dans les populations colonisatrices par rapport aux populations non colonisatrices. De plus, il semble que, plus la distance à l'origine géographique de l'espèce est importante, plus l'activité de l'élément est élevée. La mobilisation des ET a souvent été associée à la colonisation de nouvelles niches environnementales, comme la conséquence des nombreux stress (biotiques et abiotiques) auxquels sont confrontés les populations migrantes. Cependant, nous n'avons trouvé aucun lien direct entre des changements de conditions environnementales et la réactivation de *mos1* dans notre panel de 19 populations. L'étude des variations transcriptionnelles des gènes associés à la régulation par les ARN interférents, a néanmoins révélée leur implication éventuelle dans la régulation de l'activité de l'élément. Comme c'était le cas dans la lignée M19, l'expression de *mos1* n'est pas sensible aux gènes de la voie des *siARN* (*i.e.* *ago2* et *dicer2*). En revanche, l'expression de *mos1* dans la lignée germinale est négativement corrélée à l'expression des gènes *ago3* et *aubergine*, impliqués dans la régulation par les *piARN*. Enfin, l'analyse détaillée des variations d'expression du gène *Ago3* dans les 19 populations a montré que l'activité de *mos1* semblait être affectée uniquement dans les ovaires des populations colonisatrices. Ces résultats montrent que les gènes associés à la régulation des *piARN* présentent d'importantes variations. Par ailleurs, ces variations semblent affecter l'activité de *mos1* et donc, potentiellement l'activité des autres ET.

Il est intéressant de souligner que cet effet est observé uniquement chez les populations colonisatrices, qui sont caractérisées par leurs petites tailles de populations et une perte importante de diversité génétique. Nous pensons que ce dernier point est d'une importance capitale pour comprendre la dynamique des ET dans les populations naturelles. En effet lorsque les tailles efficaces de population sont grandes, les effets de la sélection naturelle tiennent un rôle essentiel, tandis que lorsqu'elles sont réduites, c'est la dérive qui est au premier plan. Notre interprétation est que dans les populations non colonisatrices, les nouvelles insertions d'ET sont rapidement éliminées par des forces de sélection purifiante. Au contraire, dans les populations colonisatrices, la sélection naturelle n'est plus (ou est moins) opérante et les nouvelles insertions peuvent être fixées aléatoirement dans les génomes, pouvant donner lieu à une augmentation du nombre de copies (comme c'est le cas pour *mos1*). Dans ce contexte, pour limiter les effets délétères associés à de nouveaux événements de transposition, la régulation *via* les *piARN*, présenterait l'avantage de limiter l'activité des ET actifs dans les populations colonisatrices. La corrélation négative entre les gènes de la voie des *piARN* et l'activité de *mos1* dans les populations colonisatrices suggère que la régulation est effective uniquement dans ces lignées.

3- La réponse du génome par les *piARN* fait suite à d'importantes vagues d'amplification

Nous avons caractérisé l'activité transcriptionnelle des familles d'ET dans les lignées germinales de espèces *D. melanogaster* et *D. simulans*. Nous avons constaté que la majorité des familles exprimées dans les lignées germinales mâles et femelles, présentaient d'importants biais liés au sexe. Par exemple, chez *D. melanogaster*, la plupart des familles d'ET sont fortement surexprimées dans les testicules. Par ailleurs, chez les deux espèces, nous avons montré que ces différences étaient principalement liées à des familles amplifiées en grand nombre de copies dans le génome. De plus, il semble que les génomes de ces deux espèces aient subi d'importantes vagues d'amplification, au moins pour certaines familles d'ET. Cependant ces « bursts » de transposition n'ont vraisemblablement pas eu lieu au même moment au cours de l'évolution des deux espèces. En effet, les principales différences que comptent les deux espèces ne sont pas liées au nombre d'insertions fixées dans le génome de référence, mais plutôt à la taille des insertions. Chez *D. melanogaster*, les familles les plus amplifiées comptent un grand nombre de copies complètes présentes à de nombreux loci dans le génome. Chez *D. simulans* les familles les plus amplifiées comptent une majorité d'éléments fragmentés dérivés de copies actives, suggérant que l'amplification massive de ces familles est d'origine ancienne.

Plusieurs banques de petits ARN ont été « mappées » sur l'ensemble des familles qui s'expriment dans les lignées germinales des deux espèces. Nous avons pu montrer que chez les deux espèces, la très grande majorité des *piARN* ciblaient préférentiellement les familles les plus amplifiées dans les génomes de références suggérant que la réponse du génome par les *piARN* est postérieure à l'amplification des familles les plus représentées. Plusieurs indices laissent à penser que la régulation par les *piARN* comporte des différences chez ces deux espèces, que l'on peut attribuer à l'activité passée ou récente des ET. Ainsi, chez *D. melanogaster*, la majorité des familles surexprimées dans les testicules, est issue de vagues d'amplification récentes et est activement réprimées dans l'ovaire. Cette situation contraste fortement avec *D. simulans* où les familles surexprimées dans les testicules sont majoritairement présentes en petit nombre de copies et peu ciblées par les *piARN*. En effet chez cette espèce les *piARN* ciblent préférentiellement de vieilles familles fragmentées dans le génome, suggérant que la réponse apportée par la voie des *piARN* est une relique issue de l'activité passée des ET. Ces résultats suggèrent finalement que la régulation par les *piARN* intervient avec un temps de latence au cours de l'invasion des ET. La régulation semble donc plus efficace lorsque l'élément a envahi le génome avec succès et qu'un pool de copies désormais inactives se soit spécialisé dans la production des *piARN* double brin.

4- La régulation par les *piARN* chez la drosophile, une spécificité de la lignée germinale femelle inopérante chez les mâles

L'étude de l'élément *mos1* dans la lignée M19 chez *D. melanogaster* et dans les populations naturelles chez *D. simulans* a permis de mettre en évidence sa surexpression au niveau transcriptionnel dans les testicules. Dans ces deux études, ce patron d'expression est en opposition avec l'expression des gènes impliqués dans la régulation par les *piARN*. L'exemple d'Ago3 montre clairement cette tendance, à la fois au niveau transcriptionnel et traductionnel dans les deux études. En effet, dans les 19 populations naturelles de *D. simulans*, alors que l'activité transcriptionnelle de l'élément *mos1* est élevée, celle d'*ago3* est particulièrement basse dans les testicules. On retrouve le même phénomène dans la lignée M19 où la forte expression de *mos1* dans les testicules corrobore la faible expression des gènes associés à la régulation par les *piARN*. Par ailleurs, dans la lignée M19, où les profils de petits ARN spécifiques de l'élément *mos1* ont pu être caractérisés, on observe une différence majeure entre les lignées germinales mâles et femelles. La quantité de *piARN* dérivés de *mos1* est de plusieurs ordres de grandeurs supérieure dans l'ovaire relativement aux testicules. Ces résultats suggèrent que la surexpression de l'élément est la conséquence d'une régulation active dans l'ovaire, peu effective dans les testicules.

Afin de savoir si ce phénomène était spécifique à notre élément d'étude ou bien une conséquence plus générale de la différenciation des deux lignées germinales, nous avons comparé les banques de petits ARN de testicules et d'ovaires chez deux espèces sœurs (*D. melanogaster* et *D. simulans*). Nous avons constatés avec surprise, des différences majeures tant au niveau quantitatif que qualitatif, directement attribuables aux différences entre les deux sexes. Premièrement, chez les deux espèces, la proportion de *piARN* (d'une taille comprise entre 23 et 29 nt) présente dans les banques issues des testicules est de 14% contre 68% dans l'ovaire. La production de *piARN* est donc beaucoup plus importante dans les ovaires comparés à celle des testicules. Deuxièmement, nous avons montré que la fraction de *piARN* ciblant les ET est beaucoup plus importante dans les ovaires que dans les testicules. Enfin, il est intéressant de constater que parmi les *piARN* dérivés d'ET, la proportion des petits ARN impliqués dans un processus de biogenèse secondaire est très largement supérieure dans l'ovaire. Grâce aux transcriptomes de testicules et d'ovaires issus de deux populations de chacune des espèces, nous avons pu quantifier les différences d'expression associées à l'ensemble des gènes essentiels au maintien de la régulation par les *piARN* dans la lignée germinale. La grande majorité de ces gènes présente un biais d'expression en faveur des femelles suggérant que ces derniers ont été favorisés au cours de l'évolution par des pressions de sélection spécifiques des femelles.

En définitive, chez *D. melanogaster* et *D. simulans*, nous avons constaté que la régulation par la voie des *piARN* semblait efficace pour réprimer l'activité des ET uniquement dans la lignée germinale femelle. Il semble que la lignée germinale mâle soit beaucoup moins contrainte à ce niveau, ce qui pourrait avoir d'importantes conséquences évolutives, par exemple dans l'émergence de nouveaux conflits intra-génomiques. Ce point sera discuté plus loin dans la discussion.

II- Le cycle de vie des ET et leur régulation au niveau des clusters de *piARN*

Un ET peut arriver dans un nouveau génome par deux moyens : soit par introgression (si les espèces sont phylogénétiquement proches), soit par transfert horizontal (si les espèces sont phylogénétiquement éloignées). Lorsque celui-ci est transmis par introgression, la moitié du génome parental est transféré avec lui. Il peut inclure plusieurs copies actives, le matériel génétique nécessaire à sa répression (*e.g.* les clusters de *piARN*), voire des facteurs épigénétiques lorsque l'élément est introduit par la mère. Étudier la mise en place de la régulation nécessite donc que l'élément soit entièrement indépendant du contexte génomique dans lequel il s'intègre. C'est pourquoi *mos1* a été introduit par transgénèse dans une espèce dépourvue de cet élément.

1- L'élément *mos1* est dans la phase stationnaire du cycle d'invasion des ET

Le cycle de vie des ET de Classe II est souvent modélisé par une courbe exponentielle (Hartl et al., 1997), assez proche de celles décrivant les cinétiques de croissance bactériennes (Figure 1). L'événement initial correspond au transfert horizontal de l'élément dans une nouvelle espèce. La courbe décrit alors, une phase d'invasion exponentielle, au cours de laquelle le nombre de copies augmente dans le génome en raison de l'activité de transposition de l'élément autonome. La seconde phase correspond à une situation d'équilibre dans laquelle le nombre de copies dans le génome demeure stable. Dans un modèle sans sélection, une interprétation serait que l'élément ait perdu sa capacité à transposer, une autre serait que l'élément transpose autant vers de nouveaux sites qu'il s'excise à partir d'anciens sites. Dans un modèle où les insertions sont considérées comme délétères, cela implique que l'élément transpose à la même fréquence que celle à laquelle il est éliminé par sélection purifiante et/ou par simple excision. Cette phase d'équilibre est suivie d'une phase de perte, dont la trajectoire évolutive est difficilement prédictible. Au cours de cette phase, les éléments autonomes sont progressivement perdus dans la population, ce qui conduit à une baisse de l'activité globale. L'élément ne

transpose donc plus ou très peu, ce qui peut mener à son extinction totale, de manière aléatoire au cours du temps (Le Rouzic and Deceliere, 2005).

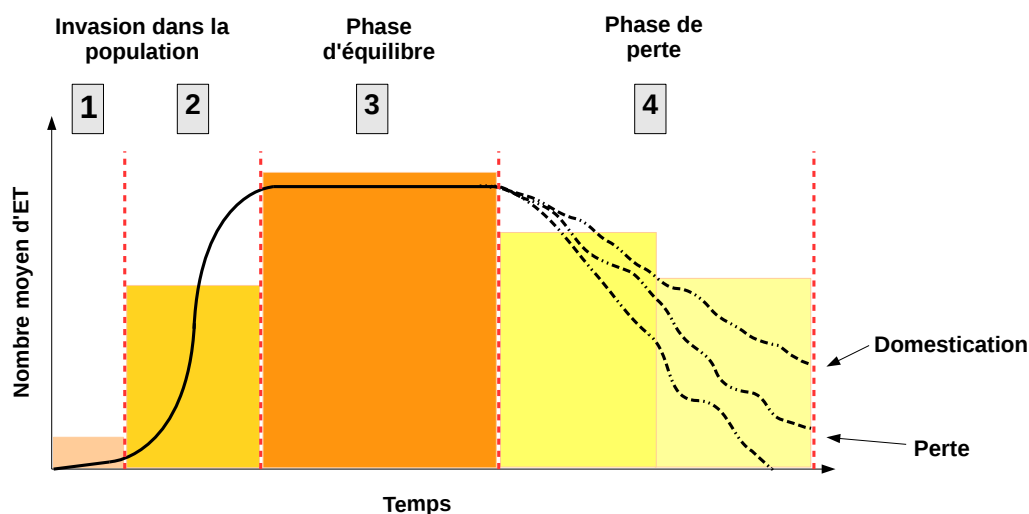


Figure 1. Dynamique d'invasion des ET de Classe II

Le graphique montre la dynamique attendue d'un ET introduit *de novo* dans une population. L'axe des abscisses représente le temps passé depuis l'insertion initiale dans la population (équivalent à un nombre de génération). L'axe des ordonnées représente le nombre moyen de copies par génome au sein de la population. L'étape 1 (orange clair) représente la première phase de l'exponentielle où l'élément commence à se diffuser dans la population. L'étape 2 (orange) représente la phase exponentielle de croissance de l'élément où l'activité de l'élément est forte. L'étape 3 (orange foncé) correspond à la phase d'équilibre où le nombre de copies est en moyenne stable dans la population (*e.g.* purgée du génome par des forces de sélection négatives et/ou par un mécanisme de régulation réprimant l'activité de transposition de l'élément). L'étape 4 (jaune) correspond à la phase de perte aléatoire de copies (*e.g.* les copies dégénèrent progressivement par accumulation de mutations et sont perdues par des forces de dérives). Eventuellement, certaines insertions sont retenues lorsque ces dernières sont domestiquées par l'hôte. Adapté de Brookfield, 2005 et (Le Rouzic et al, 2007).

Dans notre étude d'invasion de l'élément *mos1* dans la lignée M19, le nombre de copies demeure stable depuis plusieurs années, bien que l'élément soit toujours capable de se mobiliser. Si on s'en réfère au modèle décrit plus haut, l'élément serait dans une phase stationnaire, suggérant un équilibre entre gain et perte des copies. Nos observations suggèrent que cet équilibre décrit une dynamique complexe au cours de laquelle certaines insertions s'établissent de manière stable, tandis que d'autres sont régulièrement renouvelées. En effet, certaines copies semblent présentes depuis plusieurs années chez plusieurs individus, tandis que d'autres sont détectées de manière unique chez un seul individu. Par ailleurs, l'ensemble des copies n'a vraisemblablement pas le même statut du point de vue de la régulation. En effet, parmi les différentes copies identifiées, certaines sont insérées dans des régions riches en ET qui s'alignent avec un grand nombre de *piARN*. Au contraire, les insertions euchromatiques, que l'on trouve généralement dans des gènes ou à proximité, ne s'alignent qu'avec très peu de *piARN*. On peut donc supposer que dans le modèle présenté plus haut, toutes les copies ne soient pas équivalentes. Certaines seraient impliquées directement dans la production de *piARN* tandis que d'autres continueraient de produire des ARNm et par conséquent de la transposase. Dans ce modèle où l'activité de transposition diminue mais où le nombre de copies demeure constant, on peut supposer qu'il existe un équilibre entre des copies productrices de *piARN* et des copies actives dont le taux de transposition est affecté par la

régulation (Figure 2). Un tel modèle est particulièrement séduisant puisque dans cette situation, il n'est pas nécessaire d'invoquer la sélection contre de nouvelles insertions. Si par dérive une telle insertion est fixée dans un cluster de *piARN*, elle suffit à expliquer cet état d'équilibre de part son effet répresseur sur le pool de copies actives. On sait par ailleurs, qu'au cours de la vie d'un individu, une telle insertion peut arriver de manière aléatoire (Khurana et al., 2011). En effet, il a été montré expérimentalement que dans un croisement dysgénique, causé par l'élément *P*, la stérilité de la mère est restaurée au cours du temps. La réversion de ce phénotype a été associée à de nouvelles insertions de l'élément dans différents clusters de *piARN* dans les cellules germinales de quelques ovarioles qui, par ce biais, acquièrent la capacité à réprimer *de novo* l'élément *P*.

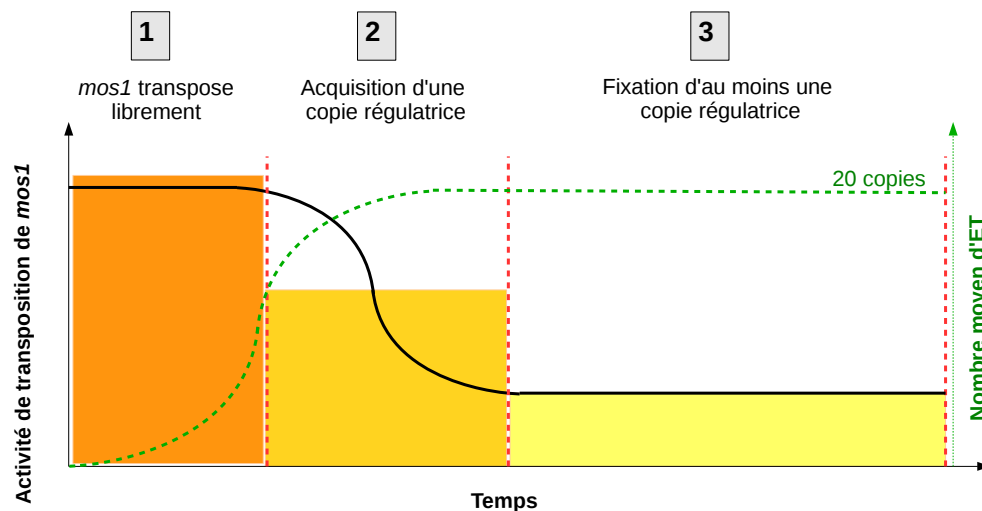


Figure 2. Dynamique de *mos1* dans la lignée M19

Le graphique est une courbe théorique qui illustre l'activité de transposition de *mos1* depuis son insertion dans la lignée M19 (courbe noire) et l'évolution du nombre de copies (courbe verte) au cours du temps (axe des abscisses). La phase 1 correspond à l'amplification de *mos1* dans la lignée, lorsque le génome M19 n'a pas encore acquis la capacité à réguler l'activité de l'élément (l'activité de l'élément est forte et le nombre de copies augmente rapidement). La phase 2 correspond à la première insertion de *mos1* au sein d'une région productrice de *piARN* chez un individu de la lignée M19 (l'activité de l'élément diminue et le nombre de copies augmente de moins en moins rapidement). Au cours de cette phase, la copie productrice de *piARN* s'amplifie dans la population et réprime progressivement l'activité de *mos1* chez tous les individus. La phase trois correspond à la fixation d'au moins une copie productrice de *piARN* dans la lignée. L'élément *mos1* est réprimé de manière stable (l'activité de *mos1* est faible et le nombre de copies n'augmente plus).

2- La dynamique des ET et leur régulation par les clusters de *piARN*

L'élément *P* compte parmi les rares cas naturellement observés de transferts horizontaux chez *D.melanogaster* (Silva et al., 2004). Pendant la phase d'invasion de cet élément, les taux de transpositions sont exceptionnellement élevés (Good et al., 1989; Kidwell et al., 1988). De plus, l'expansion de l'élément à l'ensemble de l'espèce, a eu lieu sur une courte échelle de temps, moins d'une cinquantaine d'années (Anxolabehere et al., 1988; Daniels et al., 1990). Tandis que toutes les populations collectées récemment possèdent l'élément, les lignées collectées au début du siècle en sont exemptes. Un croisement entre ces deux types de souches peut donner lieu à un syndrome de dysgenèse hybride, qui se caractérise par l'existence de

recombinaison chez les mâles, de la stérilité suite à l'atrophie des gonades et de nombreux réarrangements chromosomiques (Engels and Preston, 1979; Hiraizumi, 1971; Kidwell et al., 1977; Kidwell and Novy, 1979), imposant par conséquent, une forte pression de sélection à l'hôte. Il a été montré que certaines insertions de l'élément *P* conféraient une résistance au phénomène de dysgenèse hybride, notamment *via* des insertions dans des régions productrices de *piARN* qui répriment la transposition de l'élément. De plus, les *piARN* produits dans le cytoplasme maternel peuvent être transmis à la descendance de manière à initier le processus de répression de l'élément *P*, qui peut s'établir à travers les générations, indépendamment du site initial de production (de Vanssay et al., 2012).

Dans la lignée M19, on ignore si les insertions dans des régions productrices de *piARN* sont le fait d'insertions aléatoires maintenues par dérive, ou bien s'il s'agit d'insertions sélectionnées pour leur avantage à réprimer les autres copies actives (Figure 3). Pour *P*, il a été montré qu'une insertion ayant la capacité à réprimer la transposition de l'élément se diffusait rapidement dans les populations naturelles de *D. melanogaster* (Ronsseray et al., 1991). Il a été suggéré que cet allèle était soumis à de fortes pressions de sélection en raison de son avantage à limiter les effets délétères résultant de la transposition des autres copies (*e.g.* cassures double brin). Un modèle récent, incluant la sélection et la régulation par les *piARN* aux dynamiques classiques a été discuté autour de l'élément *P* pour lequel de nombreuses données structurales et fonctionnelles sont disponibles.

Ce modèle proposé par Lee et Langley en 2012, décrit l'évolution de l'élément *P* dans une population d'hôtes diploïdes de taille infinie initialement vide de l'élément. Leur modèle s'intéresse plus particulièrement à trois aspects de la population hôte qui change au cours des générations: (1) la proportion d'individus de cytotype *P* (résistants) dans la population; (2) la fréquence des allèles réduisant la transposition de l'élément *P* (allèle répresseur); (3) le nombre de copies moyen de *P* chez les individus de cytotype *P*. Plus précisément, ce modèle a pour but de comprendre la dynamique des allèles répresseurs au cours de l'expansion de l'élément *P* dans les populations. Il semble tout d'abord que l'acquisition d'allèles répresseurs soit directement dépendante de la vitesse à laquelle l'élément *P* se diffuse dans la population. En effet plus l'élément se diffuse rapidement, plus les risques de dysgenèse hybride sont importants. Par conséquent, une diffusion rapide conduit à l'acquisition rapide d'allèles répresseurs qui présentent un avantage sélectif. Cependant, cet avantage n'est effectif que sur une courte période de temps car, une fois que tous les individus de la population ont acquis le cytotype *P*, l'élément est réprimé *via* le cytotype maternel. Dans ce cas, la répression de l'élément est transmise indépendamment de l'allèle producteur de *piARN*. L'hôte ne présente donc plus d'avantage à fixer un allèle répresseur (Figure 4a). Dans le cas d'une invasion rapide, telle que celle observée pour l'élément *P*, on s'attend donc à ce que le nombre d'allèles répresseurs dans la population croisse rapidement, puis lentement (Figure 4b).

L'idée qu'une proportion importante d'ET soient rapidement fixée, de part leur avantage à réprimer les autres copies de la même famille, a été discuté par d'autres, de manière plus globale. Ainsi, Lu et Clark en 2010 ont proposé un modèle de dynamique supporté par des données analytiques, spécifiques de l'ensemble des

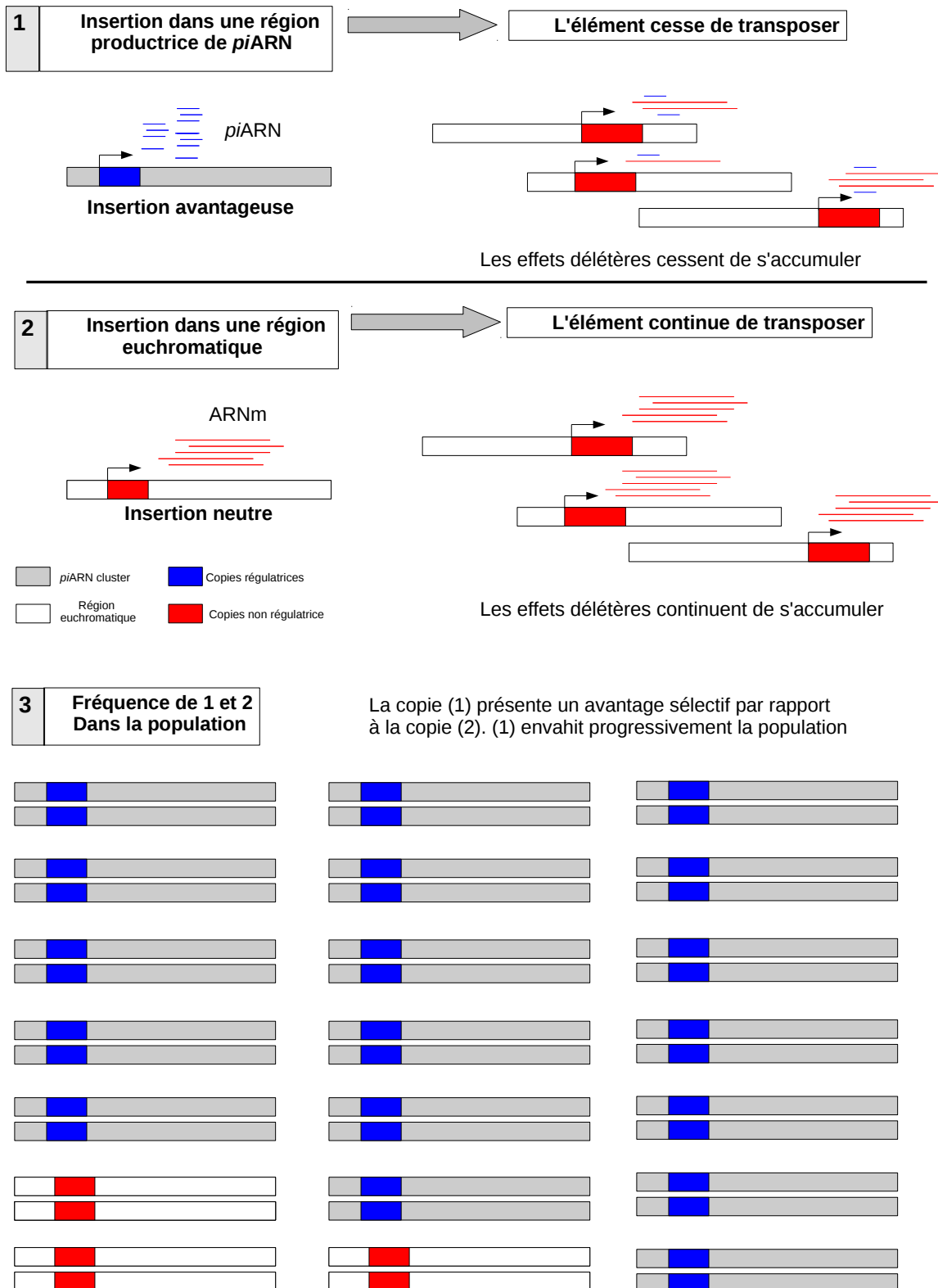


Figure 3. L'avantage sélectif des copies productrices de *piARN*

La figure schématise l'émergence d'une copie soit dans une région régulatrice (Gris) soit dans une région euchromatique (blanc). Dans le cas 1, la copie (bleu) devient productrice de *piARN* et réprime l'activité transcriptionnelle des autres copies en *trans*. Dans le cas 2, la copie (rouge) est transcrite normalement et n'a aucun effet sur les autres copies du génome. A terme (3), la copie régulatrice (1) présentant l'avantage de limiter les effets délétères des autres copies sera sélectionnée et envahira progressivement la population.

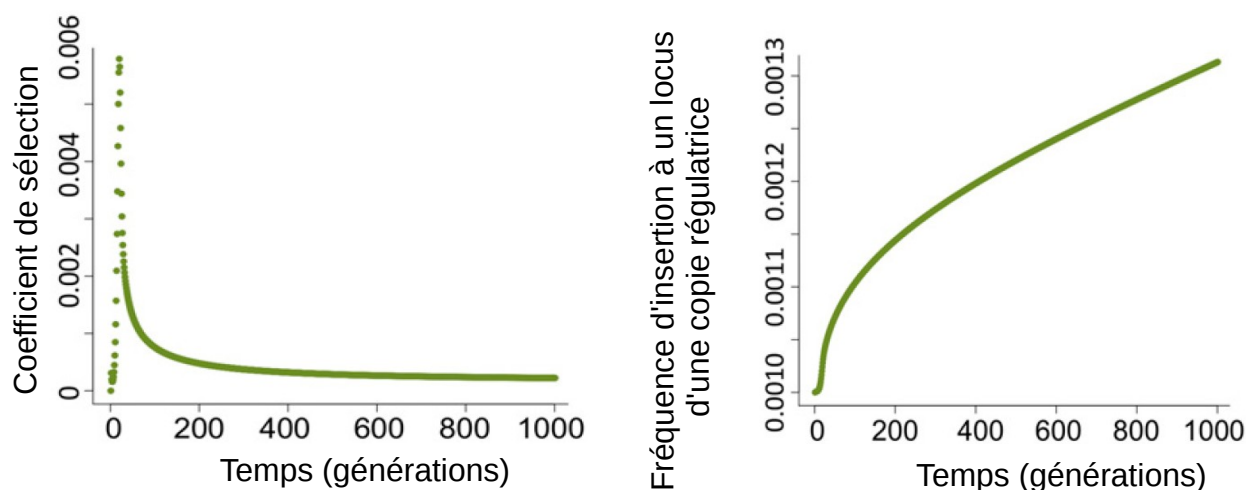


Figure 4. Dynamique d'invasion d'une copie régulatrice (sélection et fréquences d'insertions)

La figure a) représente l'évolution du coefficient de sélection (axe des ordonnées) suite au transfert horizontal d'un ET dans une population en fonction du temps (axe des abscisses). Lorsque l'ET arrive dans la population, celui-ci transpose activement et est associé à d'importants effets délétères pour l'hôte (coefficient de sélection élevé). Une fois que l'hôte a acquis la capacité à réprimer l'activité de l'élément ces effets diminuent, de même que l'activité de transposition de l'élément (coefficient de sélection faible).

La figure b) montre la vitesse de fixation d'une copie régulatrice dans la population. L'axe des ordonnées correspond à la fréquence d'une copie régulatrice à un locus dans une population, l'axe des abscisses correspond au temps exprimé en générations. Les coefficients de sélection étant élevés, la copie régulatrice augmente rapidement en fréquence dans la population, puis cette vitesse diminue lorsque les pressions de sélections sont relaxées. Ces modèles supposent une population infinie en panmixie (la probabilité d'insertion à un locus donné suit une loi de Poisson). Adapté de (Lee and Langley, 2012).

rétrotransposons. Leurs simulations indiquent premièrement, qu'en présence de *piARN* le coût sélectif imposé par les rétrotransposons est moindre, suggérant par conséquent que la régulation des ET par les *piARN* présente un net avantage pour l'hôte. Par ailleurs, pour une même famille d'éléments, il semble que les copies productrices de *piARN* soient préférentiellement acquises pour leur avantage à réprimer le pool de copies euchromatiques. Ces prédictions indiquent qu'au cours d'une dynamique d'invasion, ces copies atteignent plus facilement de fortes fréquences et peuvent être fixées. Une fois la répression établie de manière stable dans la population, les effets délétères associés à la famille diminuent. Par conséquent, relativement aux familles d'ET non réprimées, les copies ciblées par les *piARN* ont elles aussi une plus forte probabilité de se fixer dans la population. Ces simulations suggèrent qu'une fois réprimé, les ET ont de plus grandes chances de s'établir sur le long terme et que la régulation par la voie des *piARN* agit comme un « cheval de troie » facilitant l'expansion des lignées d'ET dans le génome.

En accord avec ces simulations, une autre étude (Kelleher and Barbash, 2013) souligne les limites de la voie de régulation par les *piARN* à défendre efficacement les génome contre de nouvelles invasions d'ET. Ces travaux montrent que les variations associées à l'abondance des *piARN* dans les ovaires de *D. melanogaster* s'expliquent majoritairement par trois facteurs: (1) l'abondance des transcrits messagers des ET, (2) la robustesse de la boucle d'amplification ping-pong, (3) la distribution des ET dans les clusters de *piARN*. Leurs résultats suggèrent que les *piARN* ciblent essentiellement des familles fortement transcrites dans l'ovaire qui correspondent majoritairement à d'anciennes familles sur-représentées dans les clusters de *piARN*. En effet, les

auteurs montrent que les familles en cours d'établissement, les plus actives en terme de transposition, ne sont que rarement ciblées par les *piARN*. De cette manière, ils concluent que la machinerie des *piARN* semble inefficace à défendre le génome contre de nouvelles invasions, car la mise à jour du le répertoire de *piARN* ciblant ces familles d'ET est lente.

Conceptuellement, la régulation par les *piARN* est perçue comme la voie dédiée à la protection et au maintien de l'intégrité du génome en réprimant les parasites génomiques (Siomi et al., 2011). Cependant, plusieurs auteurs tendent à penser que cette voie de régulation est globalement inefficace pour réprimer l'expansion de nouvelles familles d'ET (notamment celles retrouvées à basse fréquence dans le génome). D'autre part, ils suggèrent qu'une fois pris en charge par les *piARN*, les familles d'ET ont un avantage pour s'établir de manière stable dans le génome (être présent à haute fréquence, voire à être fixé). Dans la lignée M19, nous avons montré qu'au cours de l'invasion de *mosI*, le génome a établi sa répression par la voie des *piARN*. Ces résultats montrent clairement que cette voie agit activement suite à l'arrivée de nouveaux « parasites ». Néanmoins, les conditions expérimentales de cette étude ne sont similaires ni aux conditions décrites par les modèles, ni aux conditions naturelles dans lesquelles un ET émerge dans une population. Par exemple, les modèles considèrent l'arrivée d'un nouvel ET dans une population en panmixie, tandis que la lignée M19 est maintenue *via* des croisements frères-soeurs depuis plusieurs années. Par ailleurs, la taille de population est un facteur important qui affecte la cinétique d'invasion, notamment au niveau de la fréquence de fixation à un locus donné dans le génome. Dans notre cas, la taille efficace de population est petite, ce qui a pour conséquence de diminuer les effets de la sélection naturelle sur les nouvelles insertions. Néanmoins, nos résultats et les prédictions des modèles se recoupent en certains points, notamment *via* le recrutement d'insertions dans des régions productrices de *piARN*, qui réprime l'activité de *mosI* de telle sorte que le nombre de copies n'augmente plus dans la lignée. Cela suggère que dans des conditions où la sélection purifiante n'est pas opérante, il y a un intérêt à maintenir des allèles répresseurs pour limiter l'expansion de l'élément. Nos expériences indiquent que certaines copies sont fixées, tandis que d'autres ségrégent à plus faibles fréquences, à différents endroits du génome. Il serait intéressant de pouvoir localiser précisément l'ensemble des copies dans un grand nombre d'individus afin de pouvoir conclure quant à la proportion d'allèles de type répresseur et les autres. De plus, nous regardons l'activité de l'élément à un instant précis de la dynamique d'invasion, au niveau de la phase stationnaire. L'idéal serait de pouvoir suivre la ségrégation de l'élément depuis le point initial, dans un grand nombre de lignées pour comprendre les forces qui agissent, générations après générations, sur chacune des nouvelles insertions. Par ailleurs, il serait pertinent de reproduire cette expérience, ne serait-ce que pour savoir si la prise en charge d'un nouveau parasite par la voie des *piARN* est systématique ou non, dans quel laps de temps elle émerge et si cela passe nécessairement par une insertion dans une région déjà productrice de *piARN*.

Il semble que dans les populations naturelles, la machinerie de régulation ne cible pas particulièrement les nouvelles familles d'ET émergentes. En effet, lorsque l'on regarde les populations de *piARN* dans l'ovaire, très peu ciblent les éléments les plus actifs en termes de transposition. Ces éléments sont au début de la phase d'invasion et par conséquent, ils ne sont pas encore établis de manière stable dans le génome. Ce dernier point

est probablement déterminant pour la mise en place d'une régulation stable à l'échelle d'une population voire d'une espèce.

3- Le délai de réponse de la régulation par les *piARN*

Différents travaux se sont intéressés aux facteurs influençant la production des *piARN* sens et antisens dans l'ovaire. Sous l'hypothèse que la sélection naturelle maximise la répression des ET actifs dans la lignée germinale, on s'attend à ce que les familles d'ET les plus actives comptent parmi les plus abondantes dans les *piARN* dérivés d'ET. Cependant, ce phénomène n'est pas observé chez la drosophile (Kelleher and Barbash, 2013; Song et al., 2014), où l'abondance des *piARN* n'est pas corrélée au nombre de nouvelles insertions dans le génome. De plus, lorsque l'on s'intéresse au niveau de variation du pool de *piARN* produits dans les ovaires de lignées divergentes chez *D. melanogaster*, on ne constate que peu de variations entre souches. En effet, la très grande majorité des petits ARN produits, dérive des clusters de *piARN* qui semblent conservés en termes de localisation et de contenu (Song et al., 2014). Ces résultats suggèrent que la répression des ET est très largement dépendante du contenu en ET dans les clusters de *piARN*. Il semble donc que la répression d'une nouvelle famille à l'échelle d'une population nécessite au moins une insertion fixée dans ce type de régions. Nous avons vu auparavant que ce type d'insertion pouvait être rapidement sélectionné lorsque les effets délétères associés à la mobilisation de l'élément étaient extrêmes, comme dans le cas du syndrome de dysgenèse hybride causé par l'élément *P*. Néanmoins peu d'éléments sont associés à des phénotypes aussi sévères et par conséquent, les pressions de sélections sont moins fortes pour réprimer l'élément. Dans ce cas, on peut raisonnablement penser que l'acquisition d'une copie régulatrice mette plus de temps à s'établir à l'échelle d'une population (Figure 5).

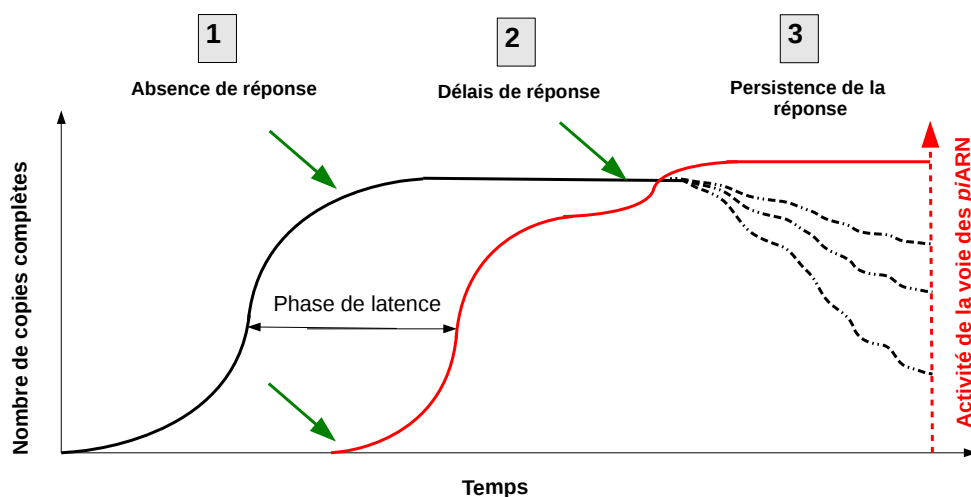


Figure 5. Latence et persistance de la réponse apportée par les *piARN* suite à une vague d'amplification

Le schéma montre l'évolution du nombre de copies (ligne noire) et l'intensité de la réponse par les *piARN* (ligne rouge) dans une population. L'étape 1 correspond à la phase d'amplification exponentielle de l'élément dans la population. L'étape 2 correspond à l'émergence d'une copie régulatrice par insertion dans un cluster de *piARN*, l'intensification de la réponse correspond à la fixation de cette copie dans la population. L'étape 3 correspond à la dégradation progressive des copies complètes, les copies séquestrées dans l'hétérochromatine produisent toujours activement des *piARN*. La flèche verte correspond à une nouvelle insertion dans une région productrice de *piARN*.

Nous avons regardé les facteurs influençant la production de *piARN* chez *D. melanogaster* et *D. simulans*. Nous avons observé chez les deux espèces que la majorité des *piARN* ciblait des éléments hautement amplifiés dans le génome (Figure 6). En revanche, les ET peu amplifiés, qui correspondent en partie à des éléments jeunes en cours d'invasion, sont globalement peu ciblés par les *piARN*. En comparant les deux espèces, nous avons pu mettre en évidence une différence majeure entre les deux génomes, liée au tempo d'activité des ET. Les familles d'ET les plus représentées chez *D. melanogaster* comptent un grand nombre d'éléments complets encore actifs, cependant associés à un certain nombre d'éléments en cours de dégradation. Chez *D. simulans*, la plupart des insertions détectées dans le génome de référence correspondent à des fragments d'ET dérivants d'élément actifs. Ces résultats suggèrent qu'une vague d'amplification est en cours chez *D. melanogaster*; tandis que chez *D. simulans* une vague d'amplification massive a eu lieu par le passé. On constate que la réponse du génome *via* les *piARN* fait suite à ces vagues d'amplifications, avec un temps de latence qui semble correspondre à l'établissement stable des lignées d'ET. De manière intéressante, la proportion de *piARN* impliquée dans la biogenèse secondaire, est nettement supérieure chez *D. simulans*, effet probablement lié à l'âge des copies. Que ce soit au niveau global ou famille-spécifique, il semble que plus les ET sont dans un état de dégradation avancé, plus la quantité de *piARN* qui en dérive est importante. Une hypothèse possible serait que ces fragments d'ET sont situés dans des clusters de *piARN* et, par ce biais, perpétuent la transcription de *piARN*, constituant en quelque sorte une relique d'une régulation ancienne (Figure 7).

D'après ces observations nous avons proposé une hypothèse associant la dynamique d'invasion des ET et la réponse associée par les *piARN*. Lorsqu'une famille active d'ET arrive dans une nouvelle espèce, celle-ci se multiplie et diffuse dans la population. A ce stade, l'élément est présent à une multitude de loci différents et la réponse par les *piARN* ne peut être effective à l'échelle de la population. En effet, même si certains individus ont acquis de nouvelles insertions dans les clusters de *piARN*, celles-ci ne sont qu'à faible fréquence dans la population. Au fur et à mesure que l'élément s'amplifie, les effets délétères associés à la mobilisation de l'élément (*e.g.* recombinaison ectopique, mutations insertionnelles) s'intensifient, en même temps que l'avantage à réprimer son activité. Par ce biais, les insertions dans les clusters vont pouvoir être sélectionnées, se fixer dans la population et ainsi réprimer de manière stable les autres copies actives. Une fois réprimée, l'activité de l'élément diminue, favorisant à terme, la fixation d'un grand nombre de copies. Sur le long terme, les copies actives fixées commencent à dégénérer, mais les fragments d'ET conservés au niveau des clusters continuent de produire une grande quantité de *piARN* jusqu'à la perte totale de l'activité de la famille (Figure 6 et 7).

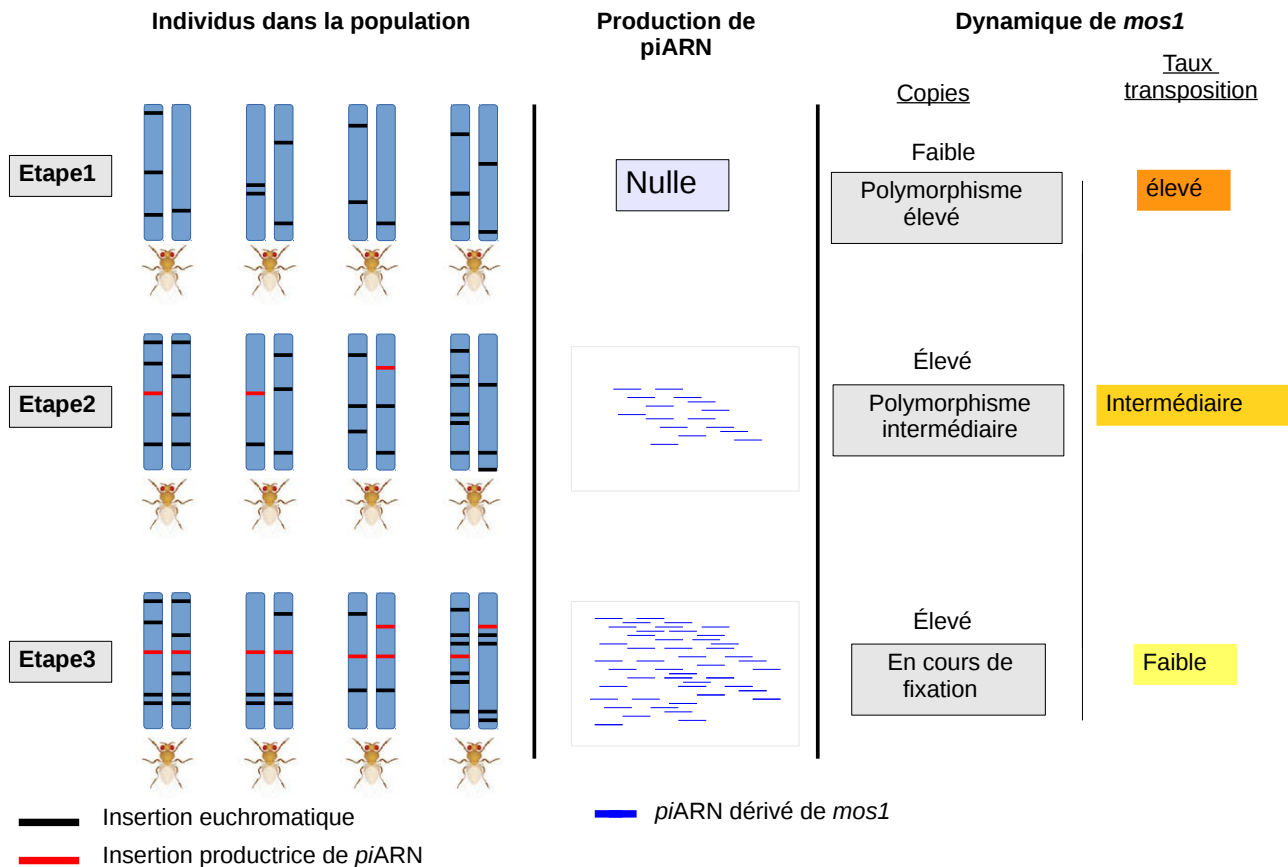


Figure 6. Établissement d'un système « silencieux » stable

Le schéma montre l'état des insertions de *mos1* dans une population. Les barres bleues schématisent le génome diploïde des individus. Les lignes noires représentent les insertions euchromatiques de *mos1*. Les lignes rouges représentent les insertions de *mos1* dans les régions productrices de *piARN*. Les lignes bleues schématisent les *piARN* dérivés de *mos1*. L'étape 1 correspond à la phase initiale d'invasion de l'élément, lorsque *mos1* s'amplifie et envahit le génome à de nombreux sites. L'étape 2 correspond à l'émergence d'un locus régulateur dans la population, qui entraîne une baisse d'activité de l'élément. L'étape 3 correspond à la phase de stabilisation de la dynamique, où *mos1* est réprimé de manière stable dans la population.

4- Les clusters de *piARN* : le cimetière des ET

Les ET réprimés par la voie des *piARN* sont potentiellement moins délétères que les autres, ce qui peut présenter un avantage pour leur établissement chez l'hôte et, sur le long terme. Il est relativement clair que les ET les plus abondants dans le génome, sont ceux pour lesquels la production de *piARN* est la plus forte et aussi les plus représentés dans les clusters de *piARN* (Kelleher et al., 2012; Malone and Hannon, 2009). De plus, nous avons vu que la production de *piARN* peut persister longtemps après l'arrêt de la vague d'amplification, même lorsque la plupart des ET sont inactifs, suggérant qu'à terme, les éléments persistants sont des fragments insérés dans des clusters de *piARN*. Une hypothèse intéressante à tester serait de savoir si au cours d'une dynamique d'invasion, les copies régulatrices insérées dans les clusters de *piARN* n'envahissent pas progressivement le génome au détriment des autres copies actives situées dans l'euchromatine (Figure 7).

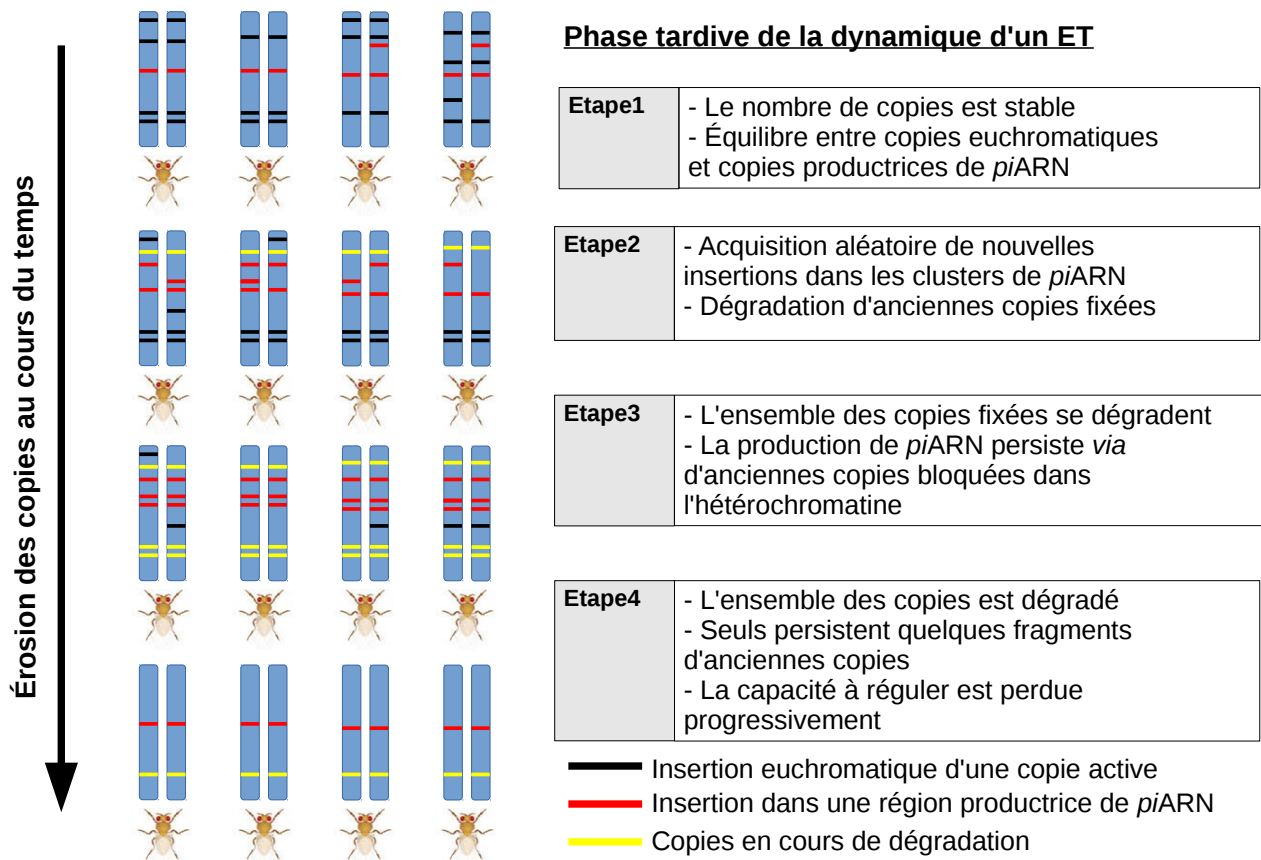


Figure 7. La phase tardive de la vie d'un ET dans le génome

Le schéma illustre la phase d'érosion du cycle de vie d'un ET depuis la phase de stabilisation. Les paires de rectangles bleus schématisent le génome diploïde d'un individu dans une population. Les lignes noires désignent les copies actives, les lignes rouges les copies productrices de *piARN* et les lignes jaunes des copies anciennes en cours de dégradation. L'étape 1, correspond à la phase d'équilibre, lorsque les copies (régulatrices et non régulatrices) sont en cours de fixation dans la population. L'étape 2 correspond à l'étape où les copies fixées commencent à dégénérer (jaune et rouge), tandis que de nouvelles insertions aléatoires sont apparues au sein de régions productrices de *piARN* et de régions euchromatiques. L'étape 3 correspond à la phase où la majorité des copies (jaune et rouge) sont fragmentées et incapable de se mobiliser. La production de *piARN* est toujours assurée par les fragments subsistant dans les clusters de *piARN*. L'étape 4 illustre la phase d'extinction de l'élément, lorsque les anciennes copies sont des reliques non reconnaissables. Au niveau des clusters de *piARN*, l'insertion continue de nouveaux ET remplace progressivement les anciens fragments de l'élément et la production de *piARN* diminue progressivement, au rythme où les fragments de copies régulatrices sont perdues.

Les copies inactivées par les *piARN* sont malgré tout associées à des effets négatifs, notamment en modifiant l'expression des gènes (Figure 8). Les *piARN* peuvent agir de manière post-transcriptionnelle en dégradant les ARNm des ET via l'activité de clivage des protéines de la famille Argonaute. Cependant, la répression peut se faire dans le noyau, au niveau transcriptionnel, par la formation d'hétérochromatine au niveau de l'insertion. L'un des effets produit par ce mécanisme, est la diffusion de l'hétérochromatine dans les régions avoisinantes, induisant leur répression transcriptionnelle. Une étude récente a montré que les gènes proches des insertions d'ET réprimés par les *piARN*, étaient associés à une baisse d'activité transcriptionnelle de la région (Lee, 2015). De plus, l'analyse génomique de ces insertions au niveau populationnel, montre que ces dernières ségrégent à basses fréquences, suggérant qu'elles sont éliminées par la sélection naturelle. Lorsque la régulation

par les *piARN* est effective à l'échelle de la population, les effets délétères associés aux copies non productrices de *piARN* persistent. Cela pourrait conduire à terme, au maintien préférentiel des copies régulatrices plutôt que des copies ciblées par la régulation.

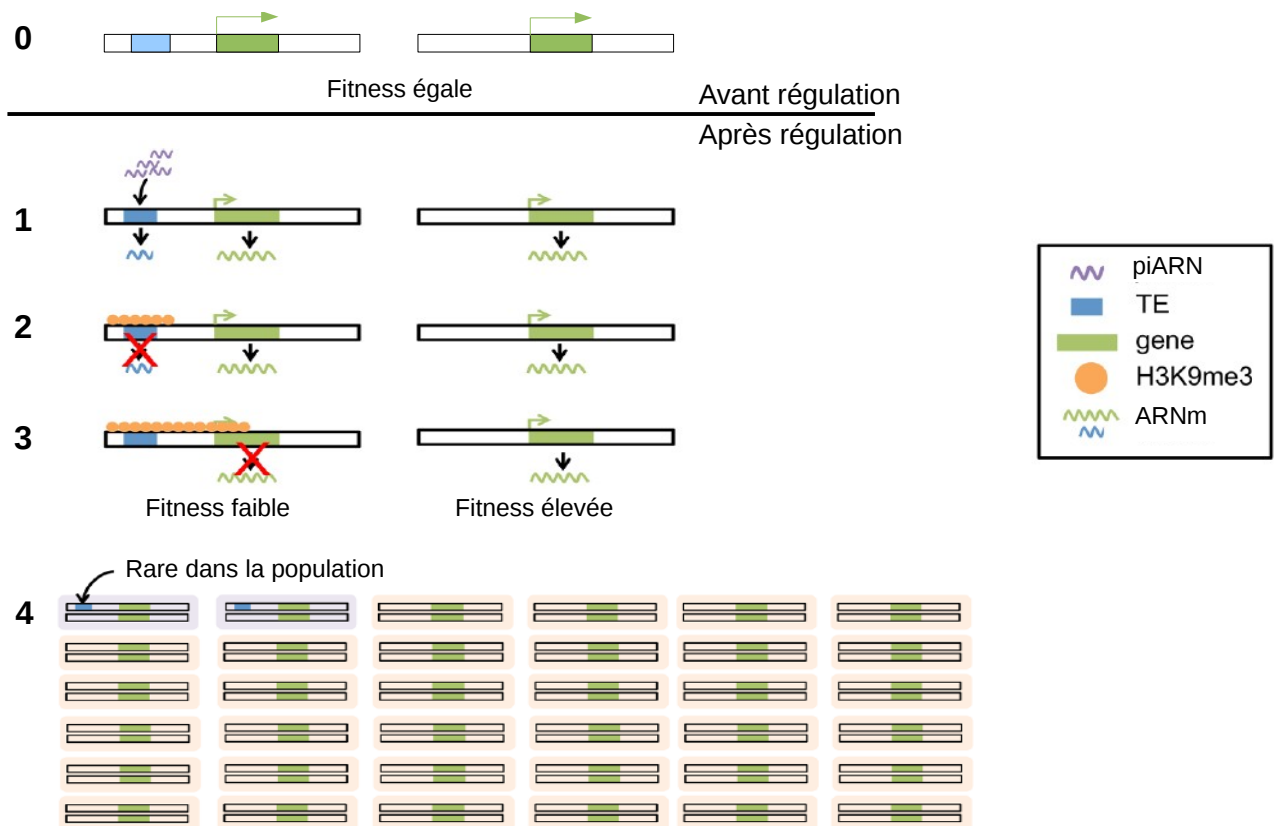


Figure 8. Effets délétères liés aux insertions d'ET soumises à la répression par la voie des *piARN*

L'insertion d'un ET à proximité d'un gène n'affecte pas l'expression de ce dernier, lorsque l'élément n'est pas réprimé par les *piARN* (0). Les *piARN* ciblent les ET qui possèdent une séquence complémentaire (1), celui-ci induit la mise en place de marques hétérochromatiques sur la séquence de l'ET (2). Les protéines se liant à l'hétérochromatine favorisent la diffusion locale des marques hétérochromatiques. La diffusion d'hétérochromatine à partir de l'ET conduit à la répression transcriptionnelle du gène adjacent (3). Les individus portant une telle insertion d'ET ont une fitness moins élevée (4 mauve), relativement aux individus sans cette insertion (4 orange). Les ET qui apportent ce type de mutations sont contre sélectionnés et sont à faibles fréquences dans la population (4). Adapté de (Lee, 2015).

Par ailleurs, il a été évoqué à plusieurs reprises que certaines insertions ont la capacité à initier *de novo* la production de *piARN* (Khurana et al., 2011; Shpiz et al., 2014; Zhang et al., 2014). L'idée proposée par ces auteurs est qu'une nouvelle insertion euchromatique peut devenir en soi, un nouveau *locus* producteur de *piARN* et induire la production de *piARN* dans la région autour de l'insertion. Néanmoins, une limite importante à cette hypothèse est que la formation de nouveaux clusters de *piARN* par une insertion d'ET dépend de l'existence d'un cluster de *piARN* déjà établi et, ciblant déjà cet ET. Les auteurs proposent qu'au cours de l'invasion, les copies actives non régulatrices ont potentiellement la capacité à devenir des copies régulatrices, suggérant qu'à terme, les nouvelles insertions deviennent des ET suppresseurs, initiant la formation des nouveaux clusters. Cette hypothèse reste cependant hautement spéculative et n'est soutenue que par quelques cas particuliers. On ignore le mécanisme sous-jacent à la formation *de novo* des clusters de *piARN*.

Au niveau évolutif, plusieurs arguments supportent que cet effet soit probablement mineur et arrive de manière transitoire au cours de l'invasion. Premièrement, la localisation des clusters est extrêmement bien conservée et l'apparition d'un nouveau cluster n'a jamais été décrite. Deuxièmement, la formation d'un cluster induit un mécanisme de transcription non-canonique en compétition avec la machinerie de transcription standard, ce qui a un impact direct sur l'expression des gènes présents dans la région (Mohn et al., 2014). Plus précisément, la formation d'un nouveau cluster peut avoir un impact négatif sur l'expression des gènes et par conséquent, va ségréger à basse fréquence dans la population (Lee, 2015). Enfin, il semble que le contenu au sein des clusters majoritaires varie énormément (Malone and Hannon, 2009), suggérant que l'adaptation à de nouveaux environnements parasitaires a lieu au niveau même des clusters de *piARN*. L'hypothèse relative à la formation *de novo* de clusters de *piARN* reste néanmoins à tester. On pourrait par exemple, comparer la synténie des clusters sur un large panel d'espèces (*e.g.* sous groupe *sophophora*) de manière à appréhender les taux d'évolution, la durée de vie et la fréquence à laquelle apparaissent ces *loci* producteurs de *piARN*.

Une question à ce jour inexplorée, est de savoir si l'acquisition des nouveaux ET dans les clusters se fait de manière aléatoire ou bien si elle résulte d'un mécanisme spécifique. En effet, les clusters de *piARN* représentent moins de 3 % de la taille totale du génome et certains d'entre eux sont composés à plus de 90 % en séquences d'ET. Cela peut être le fruit de fortes pressions de sélections pour lutter contre les effets délétères des familles d'ET actives établies dans le génome. Le temps de latence qui existe entre l'invasion d'une famille et la réponse associée par les *piARN* supporte cette hypothèse. Néanmoins, par analogie avec d'autres mécanismes de défense contre les éléments génétique mobiles, comme le système CRISPR chez les procaryotes, la question reste légitime. Les CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) sont des segments d'ADN contenant de courtes répétitions espacées entre elles par un « spacer » ADN spécifiques d'espèces virales ou de plasmides. La résistance aux éléments génétiques étrangers est acquise par l'insertion de fragments d'ADN parasites dans ces régions, à partir desquels sont produits des petits ARN. Ces petits ARN vont pouvoir guider la reconnaissance du parasite et initier sa dégradation de manière similaire aux mécanismes d'ARN interférences chez les eucaryotes. De manière intéressante, l'acquisition de fragments d'ADN étranger se fait par un mécanisme spécifique, *via* les endonucléases Cas1 et Cas2. Parmi les protéines associées à la régulation des *piARN*, un grand nombre interagissent directement avec l'ADN et ont une affinité pour les régions fortement hétérochromatiques. De plus, ces protéines ont la capacité à interagir entre elles sous diverses formes d'hétérocomplexes protéiques. Il serait intéressant de pouvoir tester *in vitro*, si ce type de complexe intervient de manière active dans l'insertion de nouveaux ET dans des séquences de type clusters de *piARN*.

III- Les conséquences évolutives de l'activité des ET dans le génome

1- ET et pathogènes : des conséquences similaires pour l'évolution des génomes ?

En permanence, les espèces sont confrontées à des changements environnementaux auxquels elles doivent s'adapter pour survivre. L'écosystème intragénomique est soumis de façon récurrente, à l'invasion de nouveaux parasites qui participent continuellement à l'émergence de nouveaux conflits. On suppose que ces conflits donnent lieu à une course aux armements entre les gènes de résistances de l'hôte et les pathogènes qui luttent chacun pour optimiser leur fitness (Figure 9). L'évolution rapide des gènes de l'immunité illustre assez bien les pressions de sélection mises en jeu dans ce type de conflit. Dans un large spectre d'espèces, les gènes impliqués dans l'immunité évoluent plus rapidement que les autres gènes. Plusieurs exemples montrent qu'ils accumulent de nouvelles mutations à des vitesses plus élevées que celle prédites par une évolution neutre. Chez *D. melanogaster* et *D. simulans*, le taux de substitution adaptative associé aux gènes de l'immunité est environ deux fois supérieur, comparé à la moyenne des autres gènes du génome (Obbard et al., 2009b). Ces résultats suggèrent que ces hôtes sont globalement mal adaptés à leur pool de parasites actuels et par conséquent, subissent de fortes pressions de sélections pour acquérir des résistances. Par analogie, en considérant les ET comme des parasites, on peut s'attendre à observer le même phénomène quant à l'évolution des gènes associés à leur répression.

En effet, chez la drosophile, les gènes antiviraux et anti-ET comptent parmi les gènes évoluant le plus rapidement. Plusieurs virus de vertébrés codent des gènes supprimant l'effet des voies de régulation par ARN interférence (Cullen, 2006). Il est donc facile d'imaginer la nature des pressions de sélection favorisant l'évolution rapide de ces gènes. En revanche, aucun ET n'est connu à ce jour comme ayant un effet inhibiteur sur les gènes associés à la régulation par les *piARN*, rendant difficile à concevoir la façon dont les ET imposent de telles pressions de sélection sur les protéines effectrices de cette voie. Une explication alternative pour expliquer ce phénomène, serait que certains effecteurs de la voie des *piARN* auraient aussi un rôle antiviral. En accord avec cette hypothèse, l'étude de la machinerie de régulation des *piARN* en contexte mutant a montré que certains virus à ARN double brin, étaient particulièrement sensibles à une déficience des ces gènes (Zambon et al., 2006). Cette idée est aussi soutenue par le fait que chez les plantes, les réponses antivirales et anti-ET sont produites par des protéines similaires (Moissiard and Voinnet, 2006). D'autre part, cette redondance fonctionnelle entre protéines anti-virus/anti-ET a aussi été suggérée chez la drosophile (Berry et al., 2009). Par ailleurs, la distinction entre rétrovirus et rétrotransposons est parfois étroite (on parle parfois de rétrovirus

endogènes en évoquant les rétrotransposons) et il est donc vraisemblable que la voie des *pi*ARN puisse supprimer l'activité de certains rétrovirus. Les protéines associées à la régulation par les *pi*ARN sont souvent associées à d'autres fonctions que la répression des ET, comme cela a été montré dans le cas de la suppression du gène égoïste *Stellate* dans les testicules de *D. melanogaster* (Aravin et al., 2007; Vagin et al., 2006). De manière générale, la plupart des gènes à évolution rapide chez la drosophile, sont ceux associés à la régulation de la conformation de la chromatine et à l'import/export de matériels nucléaires (Begun et al., 2007). Ces types de gènes sont essentiels pour le fonctionnement de la voie des *pi*ARN, mais ils sont aussi importants pour le bon déroulement de la méiose et de la reproduction. Ce dernier point peut avoir des conséquences majeures quant à l'évolution des protéines pléiotropiques impliquées dans la régulation des ET.

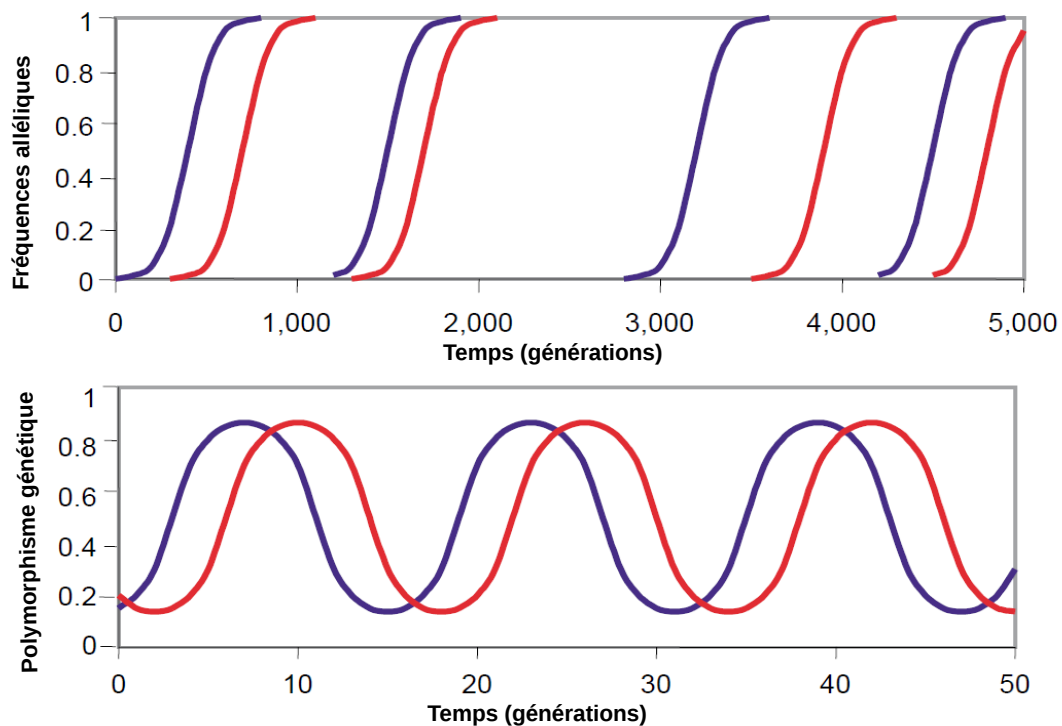


Figure 9. Les interactions hôtes-parasites : une perpétuelle course aux armements

La figure a) schématise le changement de fréquences alléliques au cours de la co-évolution entre le parasite et son hôte, à savoir une série de balayage sélectifs sur les allèles qui accumulent des mutations chez l'hôte (bleu) et chez le parasite (rouge). La sélection est supposée directionnelle, c'est à dire que les changements génétiques s'accumulent à la fois dans les populations hôtes et parasites. Sur cette figure, on voit que l'hôte s'adapte à son parasite par l'acquisition de nouveaux variants alléliques avantageux qui envahissent rapidement la population. Néanmoins, le parasite accumule aussi rapidement de nouveaux variants allélique en réponse à ces changements, quelques générations plus tard.

La figure b) illustre le polymorphisme dynamique chez l'hôte (bleu) et le parasite (rouge) agissant sur la variation génétique existante. Ici, l'évolution n'est pas directionnelle. A n'importe quelle étape, les deux populations sont polymorphes. On voit que les variations de génotype du parasite suivent de près les variations de génotype de l'hôte. On note ici aussi, un délai entre les cycles de l'hôte et du pathogène indiquant une succession permanente d'adaptations et de mal-adaptations locales pour chacune des populations. Adaptée de (Woolhouse et al., 2002).

2- L'hétérochromatine, une source de conflit favorisant la diversification des gènes associés à la régulation des ET

Les pressions de sélection pour éviter les événements de recombinaisons ectopiques induits par les ET, sont considérées comme une force majeure affectant le maintien des ET dans le génome (Kofler et al., 2012; Petrov et al., 2011). Par ailleurs, la formation d'hétérochromatine induite par les *piARN* au niveau d'une insertion d'ET, pourrait être un moyen de diminuer localement le taux de recombinaison et de limiter les effets de recombinaisons ectopiques (Peng and Karpen, 2007). Dans ce contexte, on pourrait spéculer que ces pressions de sélection affectent l'évolution des gènes associés à la répression des ET par compaction de l'hétérochromatine. Cependant les régions fortement recombinantes enrichies en ET ne sont pas particulièrement ciblées par les *piARN*, suggérant que l'hétérochromatinisation de ces régions pour limiter les événements de recombinaison ectopique est mineure (Kelleher and Barbash, 2013).

Au contraire, l'abondance des *piARN* est majoritairement expliquée par des ET insérés dans des régions faiblement recombinantes, hautement hétérochromatiques, là où l'on trouve les clusters de *piARN*. C'est probablement au sein de l'hétérochromatine, dans ce compartiment « énigmatique » du génome (Levine and Malik, 2013; Levine et al., 2012), qu'a lieu la majorité des conflits intragénomiques. Il semble que l'émergence de nouveaux conflits dans ces régions génère les pressions de sélection associées à l'évolution rapide des gènes impliqués dans la gestion et le « packaging » de l'hétérochromatine (incluant les effecteurs de la voie des *piARN*). La plupart des protéines impliquées dans la répression des ET a aussi un rôle dans la gestion de l'hétérochromatine. Une hypothèse plausible, serait que les protéines de la voie de *piARN* subissent indirectement l'effet de ces conflits, plutôt que d'être directement contraintes par l'activité des ET. Chez les eucaryotes, l'hétérochromatine se trouve essentiellement dans des compartiments structurés du génome tels que, les télomères et les centromères (20 à 30 % du génome chez la drosophile). L'assemblage de séquences et les manipulations génétiques y sont difficiles car ces régions sont hautement répétées, pauvres en gènes et fortement compactées. Ces compartiments sont le siège de nombreux processus cellulaires et évolutifs comme la régulation des gènes (Wang et al., 2000), le maintien des télomères (Cenci et al., 2005; Rong, 2008), la défense du génome (Klattenhoff et al., 2009) et la spéciation (Bayes and Malik, 2009; Brideau et al., 2006). Une étude récente a montré qu'au niveau des centromères et des télomères, les séquences répétées évoluaient extrêmement rapidement (Wei et al., 2014). L'ampleur de ces changements constitue vraisemblablement une contrainte majeure pour les protéines associées à la régulation de l'hétérochromatine (Levine and Malik, 2013; Wei et al., 2014). L'implication de ce type de protéines dans les événements de spéciation a largement contribué à l'émergence de cette hypothèse. Par exemple, pour le cas de la spéciation entre *D. melanogaster* et *D. simulans*, deux gènes (*Lhr* et *Hmr*) sont responsables de la mortalité chez l'hybride. Ces deux gènes codent des protéines interagissant avec les séquences répétées centromériques chez les deux espèces (Barbash et al., 2000; Satyaki et al., 2014). Néanmoins, ces gènes ont divergés extrêmement rapidement, de telle sorte qu'ils perdent leur capacité

à être recrutés au niveau du centromère chez l'hybride (Barbash et al., 2003; Brideau et al., 2006). Un autre type d'incompatibilité a été observé au niveau des protéines Aubergine et Argonaute3 entre ces deux espèces. Ces deux protéines, clés dans le processus d'amplification ping-pong des *piARN*, n'interagissent pas directement avec l'hétérochromatine. Elles ne sont pas présentes dans le noyau, mais à sa périphérie, au niveau du complexe *nuage*. De manière intéressante, chez les hybrides issus d'un croisement entre *D. melanogaster* et *D. simulans*, ces protéines ne sont plus localisées dans ce compartiment et il en résulte une dérégulation massive des ET. Cette incompatibilité est due à une forte divergence entre ces protéines. Un grand nombre de protéines constituant le complexe *nuage*, montre aussi des taux d'évolution rapides (Obbard et al., 2009b), suggérant que ces deux protéines se sont rapidement adaptées à ces changements pour maintenir une répression optimale des ET dans la lignée germinale. Par ailleurs, Malone *et al.* en 2009, suggèrent que les séquences présentes au niveau des clusters de *piARN* évoluent très rapidement, induisant par contre coup, une diversification rapide des protéines associées.

Les protéines impliquées dans la régulation par les *piARN* appartiennent généralement à des familles protéiques où l'on retrouve une incroyable diversité, associée, d'une part à leur caractère pléiotropique et d'autre part, aux nombreuses transitions évolutives auxquels ces protéines ont participées. Par exemple, on trouve 23 membres de la famille des protéines à domaine *TUDOR* chez *D. melanogaster*. Comme nous l'avons vu par ailleurs, un grand nombre d'entre elles ont la capacité d'interagir avec la protéine Piwi *via* des résidus arginines (*Cf* introduction). Cependant, ces protéines ont aussi la capacité d'interagir avec la queue de nombreuses histones, conduisant à leur méthylation. Plusieurs d'entre elles ont une expression spécifique dans les gonades et il a été montré qu'elles étaient indispensables au renouvellement des cellules souches et à leur maintien dans la lignée germinale. Ces protéines ont des rôles divers, associés au remodelage chromatinien et la régulation des *miARN*, *siARN* et *piARN*. Il est assez facile d'imaginer les nombreuses contraintes favorisant leur évolution rapide et leur extraordinaire diversification, sans lien direct avec l'activité des ET.

La protéine *rhino* exprimée majoritairement dans les ovaires (Vermaak et al., 2005), intervient spécifiquement dans la reconnaissance et l'activation des clusters double brin de *piARN* en partenariat avec les protéines *cutoff* et *deadlock* (Mohn et al., 2014; Zhang et al., 2014). Cette protéine appartient à la famille des protéines HP1, connue pour leurs taux d'évolution extrêmement rapide, bien que globalement conservées chez les animaux, les plantes et les champignons. Chez les eucaryotes autres que les drosophiles, la famille des protéines HP1 compte au maximum 3 membres (Vermaak and Malik, 2009). En revanche, chez les drosophiles, on en compte au moins 26 membres, illustrant l'extraordinaire diversification de ces protéines interagissant avec l'hétérochromatine. De manière intéressante, un grand nombre de protéines HP1 s'expriment spécifiquement dans les tissus germinaux et souvent, ne s'expriment spécifiquement que dans l'une des deux lignées (mâles ou femelles). L'hypothèse générale pour expliquer une diversification si rapide des protéines HP1, est qu'elles sont contraintes par les réarrangements chromosomiques et les changements au niveau des frontières entre état chromatinien et hétérochromatinien qui ont lieu au cours de l'évolution (Levine and Malik, 2013). Ces protéines sont également essentielles au cours de la méiose, où l'absence/présence d'une HP1 au niveau des

centromères peut être à l'origine des problèmes de ségrégation des chromosomes (Helleu et al., 2015). Ces deux exemples témoignent que, de manière générale, l'hétérochromatine génère de nombreux conflits évolutifs qui affectent l'évolution des protéines associées à sa gestion, notamment dans la lignée germinale.

3- La divergence entre populations naturelles : une question peu étudiée

L'étude des patrons d'expressions des gènes *aubergine* et *ago3*, effectuée dans les populations naturelles de *D. simulans*, supporte l'idée que cette diversification peut avoir eu lieu sur de très brèves échelles de temps, au sein de la lignée germinale. Par exemple, la comparaison de deux populations japonaises qui ont envahies l'île il y a moins de 50 ans, présentent parmi les variations les plus contrastées pour les gènes *aubergine* et *ago3*. Par ailleurs, nous avons observé une corrélation négative entre l'activité transcriptionnelle de ces gènes et l'activité de l'élément *mos1*. Ces résultats suggèrent donc que l'activité des gènes impliqués dans la régulation, affecte l'activité des ET et non l'inverse. Néanmoins, une étude suggère que la charge en ET dans le génome agit comme une contrainte favorisant une meilleure efficacité traductionnelle de la machinerie des *piARN*. Cette dernière hypothèse a été proposée à travers l'étude du biais de codon (plus il est fort, plus l'efficacité traductionnelle est forte) associée aux gènes de la régulation des *piARN*, cependant sans validation au niveau protéique. L'analyse des séquences codantes d'un échantillon de gènes associés à la régulation par les *piARN* chez 12 espèces de drosophiles montre une corrélation positive entre le biais de codon et l'abondance des ET (Blumenstiel, 2011). Ces résultats suggèrent que lorsque le génome contient une forte abondance d'ET, la voie de régulation des *piARN* tend à optimiser son efficacité traductionnelle pour répondre à cette augmentation. Néanmoins, ces gènes sont soumis à de fortes pressions de sélections diversifiantes et non purifiantes. Par ailleurs, le polymorphisme de séquences de ces gènes est vraisemblablement sous estimé, car, *via* l'étude des allèles présents dans le génome de référence, on occulte une grande partie des variations existantes au sein des populations de la même espèce. Une autre étude montre en effet, pour 7 gènes associés à la régulation par les *piARN*, qu'il existe un taux élevé de substitutions et que leur niveau d'expression est extrêmement variable entre populations (Fablet et al., 2014). L'étude des populations naturelles donne donc accès à un niveau de variation, à ce jour peu étudié et indispensable pour comprendre les forces évolutives auxquelles les gènes associés à la régulation par les *piARN* sont soumises. Une étude récente montre que deux lignées de *D. virilis*, présentaient des phénotypes de dysgenèse hybrides lorsqu'on les croisait (Erwin et al., 2015). Ce syndrome est associé la dérégulation de plusieurs familles d'ET qui divergent entre les deux lignées et qui sont absentes des régions péri-centriques productrices de *piARN* dans l'une des deux lignées. Ces résultats suggèrent que le contenu des régions productrices de *piARN*, au niveau des centromères, a rapidement divergé suite à l'insertion de nouveaux ET régulateurs. Dans ce cas, ces deux lignées présentent des paysages en ET très contrastés, principalement liés à l'accumulation rapide d'ET dans l'une des deux lignées. Dans ce cas, il serait intéressant de regarder l'expression des gènes associés à l'inhibition des ET pour comprendre comment ceux-ci ont été impactés (*i.e.*

l'accumulation rapide des ET a-t-elle favorisée la surexpression des gènes réprimant l'activité des ET, ou bien est-elle liée à un manque d'efficacité de la machinerie?).

Il semble clair que le pool de *piARN* produit par la mère dépend du contenu en ET des clusters de *piARN* et que les familles d'ET divergentes entre populations peuvent affecter cette production par de nouvelles insertions dans ces régions. Par ailleurs, on sait qu'entre populations, on observe des patrons d'expressions extrêmement divergents, ce qui est rarement pris en considération. Il serait intéressant de mener une étude populationnelle intégrant, le polymorphisme de séquence au niveau des clusters de *piARN* en relation avec le polymorphisme des gènes associés à leur régulation. Cela pourrait donner un aperçu direct des interactions évolutives gérées par le génome dans sa course à la répression des ET. *D. melanogaster* et *D. simulans* sont deux espèces cosmopolites pour lesquelles on peut retracer l'histoire démographique et dater leur divergence relativement facilement. On pourrait donc tester cette hypothèse *via* une étude comparative chez ces deux espèces. Plus précisément, on pourrait tester si l'évolution des protéines de la voie des *piARN* s'explique par les nouvelles invasions d'ET en cours (en ré-annotant les nouvelles insertions dans ces différents génomes), par les niveaux de divergence entre clusters de *piARN*, ou plutôt *via* des changements globaux liés au contenu hétérochromatinien (Wei et al., 2014).

Au cours de nos travaux, nous avons observé que les différences de régulation de *mos1* dans la lignée germinale des deux sexes étaient importantes. Cela suggère que le dimorphisme sexuel associé, à la fois aux ET et aux gènes impliqués dans leur répression, avait pour origine un fort antagonisme sexuel. Les conflits sexuels constituent une force majeure en évolution qui pourrait, en grande partie, être le moteur de cette diversification, c'est ce que nous allons voir dans la partie suivante.

IV- La guerre des sexes

1- Dimorphisme sexuel et sélection sexuelle

Les mâles et les femelles ont souvent des apparences et des comportements drastiquement différents. L'ensemble de ces différences phénotypiques fait référence au dimorphisme sexuel, qui résulte de la sélection naturelle ou sexuelle sur des caractères qui influencent la fitness de chacun des sexes. La sélection naturelle spécifique à un sexe, implique que cette dernière agisse en faveur de caractères associés à une survie, ou un succès reproducteur plus élevé pour un sexe que pour l'autre. La sélection sexuelle s'applique quant à elle, à des caractères spécifiquement liés au succès d'accouplement (ou de fertilisation) et sont souvent relatifs aux gènes impliqués dans la compétition sexuelle (comme par exemple la compétition spermatique ou entre mâles) mais aussi ceux relatifs à l'accouplement (comme le choix du « meilleur » mâle par les femelles).

Bien que les différences phénotypiques entre sexes puissent être extrêmes, mâles et femelles possèdent un contenu génétique quasiment identique (aux chromosomes sexuels près), voir identique chez les espèces où le déterminisme sexuel n'est pas fixé génétiquement. Ce dernier point implique que la très grande majorité des caractères sexuellement dimorphiques, résultent de l'expression différentielle de gènes présents au sein des deux sexes. Au niveau évolutif, cela sous entend que ces gènes sont soumis à différentes pressions de sélection chez les deux sexes. Cela implique éventuellement, des conflits entre les pressions de sélection qui s'exercent chez les mâles et celles qui s'exercent chez les femelles. L'antagonisme sexuel (Figure 10) décrit ce type de conflit dans lequel, un gène s'exprimant au bénéfice d'un sexe est délétère pour l'autre sexe (Figure 11). Chez la drosophile, des données expérimentales ont confirmé que la présence d'allèles sexuellement antagonistes est fréquente et que ceux-ci répondent à la sélection (Chippindale and Rice, 2001; Rice, 1996; Rice, 1998). Cela présuppose par ailleurs, que les différences d'expressions entre sexes seraient un moyen de répondre à l'émergence de nouveaux conflits sexuels et par conséquent, favoriseraient l'évolution du dimorphisme sexuel.

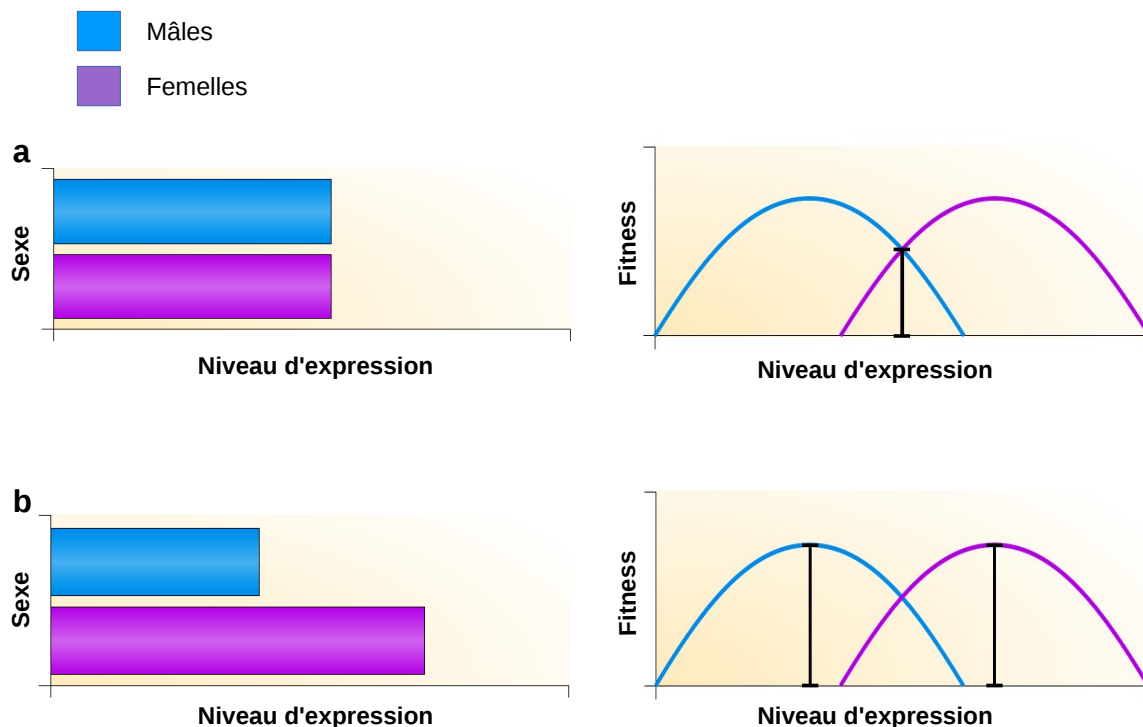


Figure 10. Antagonisme sexuel

Le panneau de gauche montre les niveaux d'expressions de gènes chez les mâles et chez les femelles lorsque l'expression n'est pas biaisée entre mâles et femelles (a) et lorsque l'expression est biaisée selon le sexe (b). Le panneau de droite montre les distributions de fitness dans chacun des sexes en fonction du niveau d'expression des gènes. Lorsque les niveaux d'expression ne sont pas différents, la fitness des individus est sub-optimale. Lorsque l'expression des gènes est biaisée selon le sexe, chaque sexe est plus proche de sa fitness optimale. Adapté de (Ellegren and Parsh, 2013).

Les gènes sexuellement dimorphiques font communément référence aux gènes « sexe-biaisés », ce qui inclut les gènes exclusivement exprimés chez l'un des deux sexes (« sexe spécifique ») de même que les gènes dont les niveaux d'expression sont plus importants chez un sexe (« sexe enrichis »). Parmi les gènes biaisés en fonction du sexe, on peut faire la distinction entre les gènes « mâles-biaisés » et ceux « femelles-biaisés » selon qu'ils soient surexprimés chez les mâles ou les femelles. Ainsi, il a été montré à travers un large spectre de taxons (nématodes, insectes, mammifères, oiseaux), que les différences d'expression liées au sexe sont intenses. L'idée que les gènes sexe-biaisés résultent de pressions de sélection sexuellement antagonistes est admise, néanmoins cela ne signifie pas que l'ensemble des gènes sexe-biaisés soient ou aient été soumis à ce type de pressions. En effet, il a été montré pour des lignées hémiclones chez *D. melanogaster* (lignées haploïdes qui peuvent s'exprimer soit comme des mâles soit comme des femelles), que seulement 8 % des gènes seraient soumis à des pressions de sélection sexuellement antagonistes. Cela contraste fortement avec le nombre total de gènes différentiellement exprimés (90%) entre les deux sexes chez cette espèce et nuance l'hypothèse selon laquelle, un gène sexuellement biaisé est nécessairement impliqué dans un conflit sexuel antagoniste (Innocenti and Morrow, 2010a; Innocenti and Morrow, 2010b).

2- Les gènes « sexe-biaisés » sont très répandus dans la lignée germinale et évoluent rapidement

L'analyse des gènes sexe-biaisés, à l'échelle du génome, a été effectuée dans un grand nombre de travaux (Ingleby et al., 2015). Il n'y a pas de consensus général pour déterminer jusqu'à quel point les transcriptomes sont sexuellement biaisés, comme le montre les larges variations entre espèces. Par exemple, chez l'escargot *littorina saxatilis*, seulement 2 % des gènes présentent un biais (Martinez-Fernandez et al., 2010), tandis que chez *D. melanogaster*, plusieurs auteurs montrent que plus de 90 % des gènes sont biaisés selon le sexe (Ayroles et al., 2009; Innocenti and Morrow, 2010a). Par ailleurs, selon les études menées chez la même espèce, le pourcentage de gènes biaisés peut varier de plusieurs ordres de grandeur (e.g. 57 % de gènes « sexe-biaisés » dans une autre étude menée chez *D. melanogaster* (Ranz et al., 2003). Outre les différences liées au type de séquençage ou à l'analyse statistique, une source majeure de variations concerne le nombre et la nature des tissus analysés ((Yang et al., 2006) pour une étude détaillée chez la souris).

Chez *Bombyx mori*, les pourcentages de gènes biaisés selon le sexe varient de 2 à 30 % suivant le type de tissu examiné (Xia et al., 2007). Chez l'espèce *Danio rerio*, seulement 2 % des gènes sont biaisés au niveau du cerveau (Santos et al., 2008), tandis que ce pourcentage s'élève à 38 % lorsque les transcriptomes des tissus reproducteurs sont comparés (Small et al., 2009). Dans les espèces du genre *Xenopus*, 70 % des gènes montrent des différences significatives entre gonades. Ces résultats illustrent une hypothèse basée sur l'étude de transcriptomes d'espèces modèles montrant que la majorité des gènes biaisés correspondent à des différences

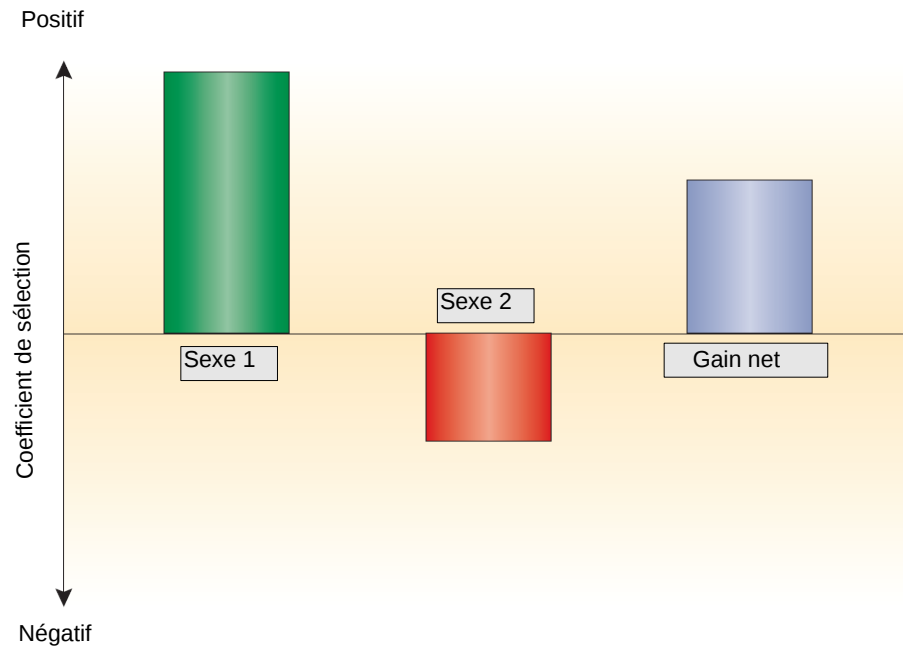


Figure 10bis. Compromis évolutif associé aux allèles antagonistes

Lorsque qu'elle est autosomale, une mutation sexuellement antagoniste peut se fixer uniquement si les bénéfices de fitness induits par la surexpression dans un des sexes sont globalement plus élevés que les effets délétères causés dans l'autre sexe. Adapté de (Ellegren, 2007).

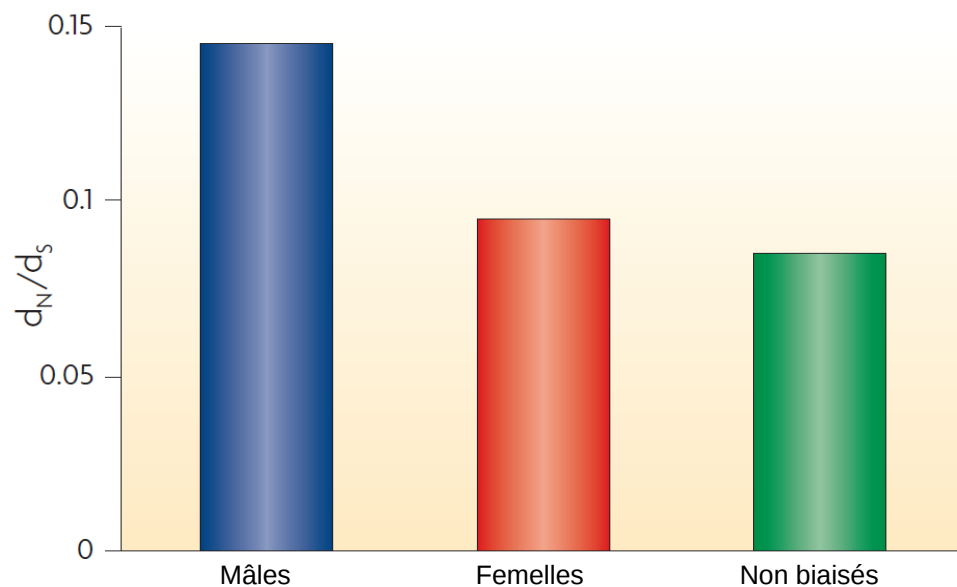


Figure 11. Divergence des gènes biaisés selon le sexe entre espèces de drosophile

Le graphique montre les ratios de substitutions non-synonymes et synonymes (d_N/d_S) pour une comparaison sur tout le génome de tous les gènes orthologues chez deux espèces proches (*D. melanogaster* et *D. simulans*). Les gènes biaisés chez les mâles montrent des taux d'évolutions plus rapides. Adapté de (Ellegren, 2007).

d'expressions au sein des gonades et des tissus impliqués dans la reproduction (Ellegren et al., 2007; Parisi et al., 2003; Parsch and Ellegren, 2013).

Par ailleurs, plusieurs études montrent que chez l'adulte, les gènes biaisés chez les mâles sont plus nombreux comparés aux gènes biaisés chez les femelles (Ellegren and Parsch, 2007; Graveley et al., 2011; Martins et al., 2013). De plus, les gènes biaisés chez les mâles montrent d'une part, des taux de renouvellement plus fréquents, d'autre part, des taux de divergence plus élevés comparé à ceux des gènes surexprimés chez les femelles ou non différentiellement exprimés (Figure 11). Ces observations sont d'autant plus fortes lorsque l'on s'intéresse uniquement aux gènes présentant une expression spécifique dans la lignée germinale. Par exemple, la comparaison entre *D. melanogaster* et *D. pseudobscura* (25 millions d'années de divergence) montre que les gènes surexprimés chez les mâles ne contiennent qu'une faible fraction de gènes orthologues identifiables entre ces deux espèces (Richards et al., 2005). L'évolution rapide des gènes biaisés selon le sexe a aussi été décrite chez les nématodes du genre *Caenorabditis*. Il s'agit d'un modèle d'étude intéressant pour la question des gènes à expression sexe-spécifiques, puisque l'on peut comparer les individus hermaphrodites aux individus mâles uniquement. L'analyse détaillée des transcriptomes (Reinke et al., 2004) de *C. elegans* et *C. briggsae* (Cutter and Ward, 2005) a permis d'établir une connexion entre patrons d'expression et taux d'évolution moléculaires. Généralement, les gènes associés à la spermatogenèse ont les taux d'évolution les plus rapides. Lorsque l'on compare les gènes de la spermatogenèse entre individus hermaphrodites et mâles exclusivement, ces derniers présentent des taux de substitution plus élevés encore. Ce constat reste vrai si la comparaison porte sur ces gènes et ceux exprimés exclusivement dans l'ovocyte. Comme chez les drosophiles, les gènes spécifiquement exprimés dans le sperme ou dans la lignée germinale mâle comptent un excès significatif de gènes orphelins (les gènes sans orthologues dans d'autres espèces).

Des études de génomique comparatives menées chez les mammifères présentent des résultats similaires. Par exemple, une comparaison entre gènes orthologues chez l'humain et la souris montre que les gènes spécifiquement exprimés dans les spermatozoïdes, ont des taux de substitution deux fois supérieurs comparés à ceux des gènes exprimés dans d'autres tissus (Torgerson et al., 2002). L'étude comparative entre génomes et transcriptomes de rats et de souris (Good and Nachman, 2005; Schultz et al., 2003) montre que les gènes exprimés tôt dans le processus de spermatogenèse présentent un taux de substitution légèrement supérieur à la moyenne des autres gènes. Ces observations contrastent fortement avec les gènes exprimés tardivement au cours de ce processus, dont les taux de substitution non-synonymes sont deux fois supérieurs à ceux des autres gènes. Entre humains et chimpanzés, les gènes spécifiquement exprimés dans les testicules présentent les mêmes caractéristiques (Khaitovich et al., 2005).

L'ensemble de ces observations a conduit à une hypothèse selon laquelle, l'évolution des gènes différentiellement exprimés entre les sexes était la conséquence de pressions de sélection spécifiques chez les mâles, comme c'est le cas pour de nombreux conflits ayant lieu dans cette lignée germinale (Artieri and Singh, 2010; Connallon and Knowles, 2005; Harrison et al., 2015; Small et al., 2009; Swanson and Vacquier, 2002).

3- Le sexe : une force motrice dans l'évolution des gènes effecteurs de la voie des *piARN*

Au cours de nos travaux, il a été observé d'une part, que les ET avaient tendance à être surexprimés dans la lignée germinale mâle, et d'autre part, que les gènes associés à la régulation des *piARN* étaient majoritairement surexprimés dans la lignée germinale femelle. Chez les mammifères, les plantes (Wu et al., 2014) et les drosophiles, plusieurs études montrent cette tendance à la surexpression dans la lignée germinale mâle (*Cf* introduction). Chez les mammifères et *Arabidopsis thaliana*, il semble que ce phénomène soit lié à un état relâché de la chromatine à certains stades de différenciation de la lignée germinale mâle (Kleene, 2001; Wu et al., 2014), qui a pour conséquence une hyperactivation de l'état transcriptionnel. La suractivation des ET dans la lignée germinale mâle peut donc constituer une source majeure de conflits délétères pour les deux sexes. Ainsi, nous avons pu montrer chez *D. melanogaster* et *D. simulans* que la machinerie de régulation par les *piARN* était associée à un fort dimorphisme sexuel chez plusieurs espèces. Une interprétation de ces résultats serait qu'en réponse à la suractivation des ET dans la lignée germinale mâle, la sélection aurait favorisé la surexpression des gènes associé à la répression des ET chez la femelle. Premièrement, la machinerie de régulation par les *piARN* constitue un coût métabolique important ce qui peut avoir favorisé sa restriction à un seul des deux sexes. Deuxièmement, le matériel cytoplasmique transmis à la descendance se fait essentiellement par la mère chez la drosophile et il est indispensable au développement de l'embryon, ce qui peut expliquer pourquoi l'investissement dans la régulation a pu être sélectionné chez les femelles. Enfin, comme nous allons le voir dans la section suivante, les perpétuels conflits au sein de la lignée germinale mâle sont favorables à l'émergence de diversité génétique, ce qui peut être induit par l'activité des ET dans cette lignée.

V- Les modèles « male driven evolution » et « out of testes »

1- L'évolution facilitée par la lignée germinale mâle (« male driven » evolution)

A - Le nombre de réplifications dans la lignée germinale mâle

Bien que la plupart des anomalies génétiques soient transmises *via* des mutations pré-existantes, une fraction importante des mutations délétères a pour origine de nouvelles mutations générées dans la lignée germinale. Bien que rares, certaines de ces mutations peuvent être importantes pour générer la diversité génétique nécessaire au processus d'adaptation. Au sein de la lignée germinale, les mutations peuvent avoir des causes endogènes ou bien exogènes, tant chez les mâles que chez les femelles. Chez l'homme, le nombre de cycles de réplication des spermatogonies étant plus élevés et les gamètes produites tout le long de la vie de l'individu (Figure 12), il a été supposé que la lignée germinale mâle avait une plus forte tendance à générer des mutations spontanées (Figure 12) (Haldane, 1947). Initialement, le taux de mutations mâles-femelles montrait une corrélation positive avec le nombre de divisions cellulaires dans chacune des lignées (Hurst and Ellegren, 1998). L'âge auquel les gamètes sont produits était donc supposé affecter ce ratio en faveur des mâles puisque chez la femelle, le nombre total de divisions reste constant au cours du temps, tandis qu'il augmente linéairement avec l'âge chez les mâles. Étant donné que la différence entre nombre de réplifications chromosomiques augmente entre mâles et femelles, la proportion de substitutions transmises par les mâles est supposée augmenter aussi. Chez l'homme, une étude menée sur 78 individus a estimé qu'en moyenne 60 nouvelles mutations apparaissaient à la génération suivante (Kong et al., 2012). Parmi ces mutations, il a été montré que 75 % ont été transmises par les pères et que l'âge du père affectait drastiquement cette proportion (2 mutations supplémentaires par années). D'autres types de mutations, comme les insertions et les délétions sont plus fréquemment transmises par les mâles, comme cela a été montré chez les rongeurs (Makova *et al.* 2004). L'analyse des génomes de plusieurs espèces tend généralement à confirmer de manière générale, que le taux de mutation est plus important dans la lignée germinale mâle et pas nécessairement lié au simple nombre de cycle de réplication (Wilson Sayres and Makova, 2011).

Il existe d'autres sources de mutations que celles liées à la machinerie de réplication. Par exemple les dommages générés par le stress oxydatif, lié à l'accumulation des dérivés réactifs de l'oxygène (ROS) dans certains tissus. Il a été montré que les cellules de la lignée germinale mâle évoluent dans un environnement riche

en ROS comparé à celui de l'oeuf, qui de plus possède une membrane cellulaire dense le protégeant d'éventuels dommages exogènes (Veland et al., 2008). Aussi, chez les mammifères, le taux de conversion spontanée des cytosines méthylées en thymine au niveau des îlots CpG, est différent dans la lignée germinale mâle. Il a donc été supposé que ce mécanisme pouvait induire une certaine hétérogénéité au niveau de ces sites (Taylor et al., 2006).

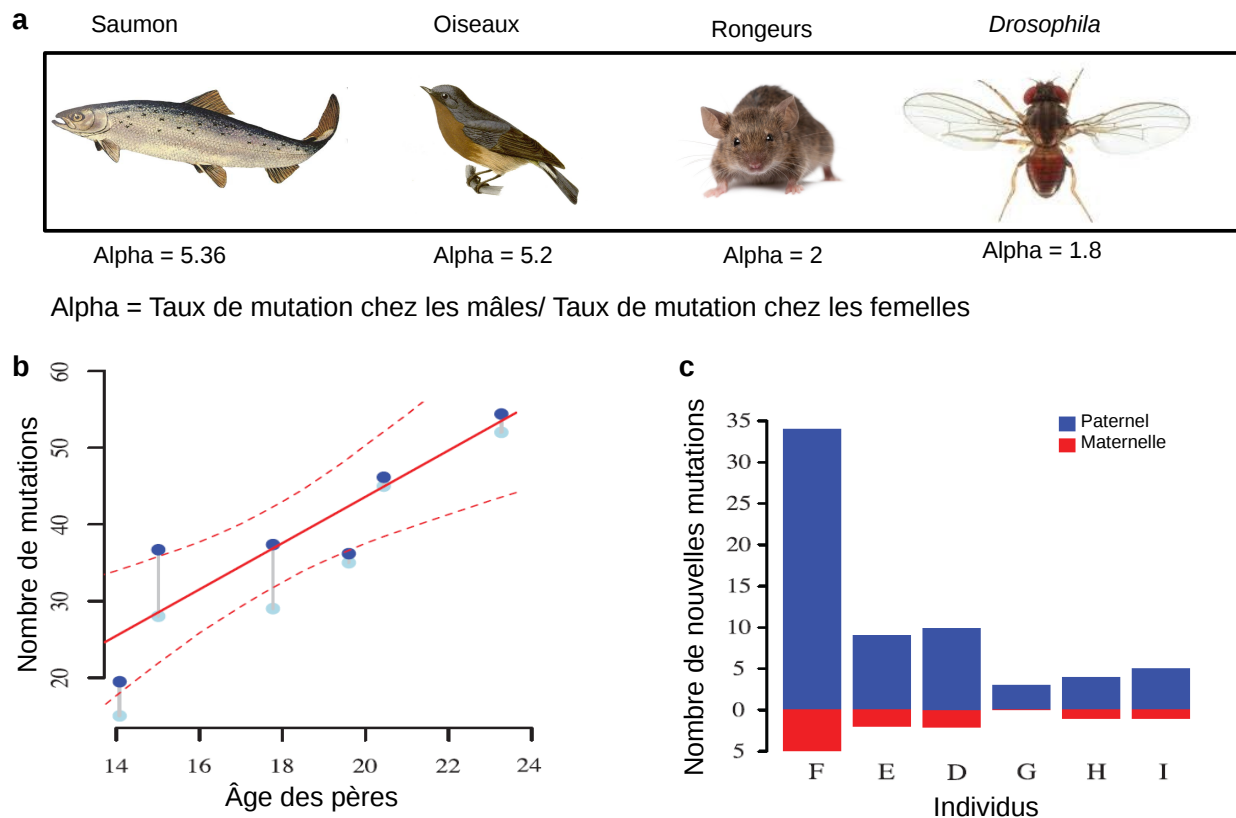


Figure 12. Le taux de mutations est plus élevé chez les mâles

La figure (a) montre les ratios entre taux de mutations chez les mâles et taux de mutations chez les femelles (alpha) chez plusieurs espèces de saumon (Ellegren and Fridolfsson, 2003), d'oiseaux (Ellegren and Fridolfsson, 1997), de rongeurs (Chang et al., 1994) et de drosophiles (Bachtrog, 2008). La figure (b) montre le nombre moyen de mutations acquises dans le sperme de chimpanzés (ordonnées) en fonction de l'âge du père (abscisses). Lorsque l'âge du père augmente le nombre de répliquations de cellules de la spermatogenèse augmente et accumule plus de nouvelles mutations. La figure (c) montre, toujours chez les chimpanzés, le nombre de nouvelles mutations acquises par les mâles (bleu) et les femelles (rouge) chez plusieurs individus (F, E, D, G, H, I). Le nombre de nouvelles mutations héritées par le père est systématiquement supérieur relativement aux femelles. Adapté de (Venn et al, 2014).

B - Le remodelage hétérochromatinien au cours de la spermatogenèse

Le processus de spermiogenèse (spermatogenèse post-méiotique), représente vraisemblablement une étape ayant une contribution importante à l'augmentation du taux de mutation plus élevé chez les mâles. En effet, durant la spermiogenèse, la spermatide haploïde s'engage dans un programme de différenciation morphologique majeur, au cours duquel les structures chromatiques et hétérochromatiques sont entièrement remodelées. La plupart des histones sont remplacées par des protamines de manière à conférer une plus grande

stabilité à la chromatine (mécanique et chimique) dans le sperme mature (Ward, 2011). Le mécanisme conduisant à une transition nucléaire aussi considérable, est à ce jour peu connu mais implique des changements post-traductionnels majeurs, ainsi qu'une dégradation générale des histones (Awe and Renkawitz-Pohl, 2010; Grégoire, 2011). Un autre fait marquant, est l'observation de nombreuses cassures double brin transitoires au cours des étapes de remodelage, notamment dans toutes les populations de spermatides chez l'homme et la souris (Marcon and Boissonneault, 2004). L'origine de ces cassures n'est pas claire à ce jour, mais pourrait résulter de mécanismes divers, soit enzymatiques (Topoisomérase de type II – McPherson et Longo, 1993), soit liés à l'activité accrue des ROS (Sakkas and Alvarez, 2010) ou simplement à des cassures mécaniques induites par des changements de structures hétérochromatiniennes (Sakkas and Alvarez, 2010). L'absence de chromatides sœurs dans les spermatides implique nécessairement des mécanismes de réparation non homologues (NHEJ), souvent associé à l'insertion ou la délétion des courtes séquences au site de réparation (Lieber and Wilson, 2010). Les cassures double brin sont aussi observées dans les spermatides de nombreuses d'espèces non apparentées, comme la drosophile (Rathke et al., 2007), la sauterelle (Cabrero et al., 2007), les rats (Meyer-Ficca et al., 2005) et même chez certaines les algues (Wojtczak et al., 2008). L'ensemble de ces observations suggèrent que la spermiogénèse est un processus physiologique extrêmement bien conservé qui pourrait avoir un impact direct sur l'évolution biaisée par la lignée germinale mâle. Par ailleurs, on peut suspecter que ce type d'événements puisse être favorable à l'apparition de nouveautés génétiques et éventuellement de nouvelles fonctions.

2- Les testicules, une niche favorisant l'émergence de nouvelles fonctions **(« out of testes hypothesis »)**

L'apparition de nouveaux gènes fournit le matériel nécessaire à l'acquisition de nouvelles fonctions moléculaires. De petites modifications au niveau d'un gène pré-existant peut donner lieu à des variations d'activités transcriptionnelles ou traductionnelles. Cependant, les nouveaux gènes possédant de nouvelles fonctions ont une contribution significative à l'évolution de caractères phénotypiques, lignée ou espèce spécifiques. Dès les années 30 (Haldane, 1933 et Muller 1935), il a été proposé que de nouvelles fonctions géniques pouvaient émerger à partir du remaniement d'anciennes copies de gènes. Aujourd'hui il est montré que les nouveaux gènes, acquis par la duplication d'anciens gènes, sont abondants dans les génomes eucaryotes et souvent associés à l'acquisition de nouvelles fonctions (Lynch, 2007; Romero and Palacios, 1997). Cependant, il existe un grand nombre d'autres mécanismes par lesquels de nouvelles fonctions peuvent émerger (Kaessmann, 2010), notamment ceux liés à l'activité des ET.

En effet, les rétrotransposons produisent des enzymes (notamment la reverse transcriptase) qui ont la capacité de rétro-transcrire l'ARN messager issu d'un gène en son ADN complémentaire et de l'insérer à un autre locus dans le génome (par une activité de type intégrase). Le gène résultant d'une rétrotransposition est inséré

sans intron et ne contient donc que l'information « exonique » du gène parental. Il est donc potentiellement impliqué dans le codage d'informations non redondantes en comparaison à une duplication classique. Ce phénomène contribue de façon significative à l'apparition de nouveaux gènes et de nouvelles fonctions (Kaessmann et al., 2009), qui de plus semble relativement fréquent (Vinckenbosch et al., 2006).

D'autres mécanismes associés aux ET semblent favorables à l'émergence de nouveaux gènes, comme les insertions domestiquées, la transduction, la fusion de gènes ou la résurrection d'anciens transposons. Bien que l'éventail de possibilités paraissent considérables, il n'y a à ce jour, pas d'étude exhaustive quant à la proportion de gènes émergents, directement attribuables à l'action des ET.

Néanmoins, un consensus semble émerger lorsque l'ensemble des études associées à l'analyse des gènes orphelins est pris en considération. En effet, dès les premières investigations sur l'origine des nouveaux gènes menées chez les mammifères et la drosophile (Long and Langley, 1993; McCarrey and Thomas, 1987), il a été montré que ces gènes étaient spécifiquement exprimés dans un tissu: les testicules. Plus tard des études focalisées sur les événements de rétroduplications ont montré une tendance générale à la surexpression des gènes orphelins dans cet organe (Betran et al., 2002; Marques et al., 2005). Sur la base de ces observations, il a été suggéré que les testicules représentaient une niche favorable quant à l'évolution des nouveaux gènes, qui par la suite pourrait acquérir des fonctions dans d'autres tissus (somatiques) avec le temps (Kaessmann et al., 2009; Marques et al., 2005; Vinckenbosch et al., 2006). Un certain nombre d'études (Kaessmann, 2010) pour une revue) a par la suite alimenté cette hypothèse, montrant que l'ensemble de nouveaux gènes présentait cette tendance à la surexpression testiculaire (les gènes issus de duplication segmentales, les gènes chimères *etc.*). Néanmoins cette hypothèse reste à tester de manière exhaustive (Figure 13).

Plusieurs facteurs supposent indirectement que l'émergence de nouveaux gènes soit facilitée dans les testicules. Premièrement, comme nous l'avons évoqué précédemment, ce tissu constitue un lieu où l'évolution peut être rapide en raison des pressions de sélection associées à la compétition spermatique, des conflits sexuels, des pathogènes de la lignée germinale, de la distorsion de ségrégation *etc* (Nielsen et al., 2005). Il semble donc que les testicules représentent un tissu impliqué dans de nombreux conflits et où l'apparition de nouveautés génétiques, peut permettre leur résolution. Une question importante reste cependant : quels mécanismes permettraient l'émergence de nouvelles ORF spécifiquement transcrites dans les testicules. Une réponse possible est apportée par l'étude des propriétés transcriptionnelles des cellules méiotiques et post-méiotiques chez les mammifères. En effet, plusieurs travaux montrent qu'à ces stades, plusieurs variants spécifiques d'histones et de modifications semblent favoriser la conformation de la chromatine dans un état « ouvert » (Kimmins and Sassone-Corsi, 2005; Kleene, 2001; Sassone-Corsi, 2002). En association avec la déméthylation globale des promoteurs enrichis en îlots CpG et la suractivation de certains composants de la machinerie transcriptionnelle (Kleene, 2001), ces facteurs pourraient conduire à un état chromatinien, permissif à la transcription *de novo*, impliquant par conséquent la transcription d'éléments non-fonctionnels ou bien habituellement réprimés. Un sous-échantillon de ces éléments constitue une source de matériels potentiellement sélectionnables, qui à terme,

pourrait être recrutés pour des fonctions avantageuses dans d'autres tissus. Néanmoins cela implique qu'une grande partie de ces gènes soit restreinte à la lignée germinale et par ailleurs, rapidement contre-sélectionnée pour la majorité.

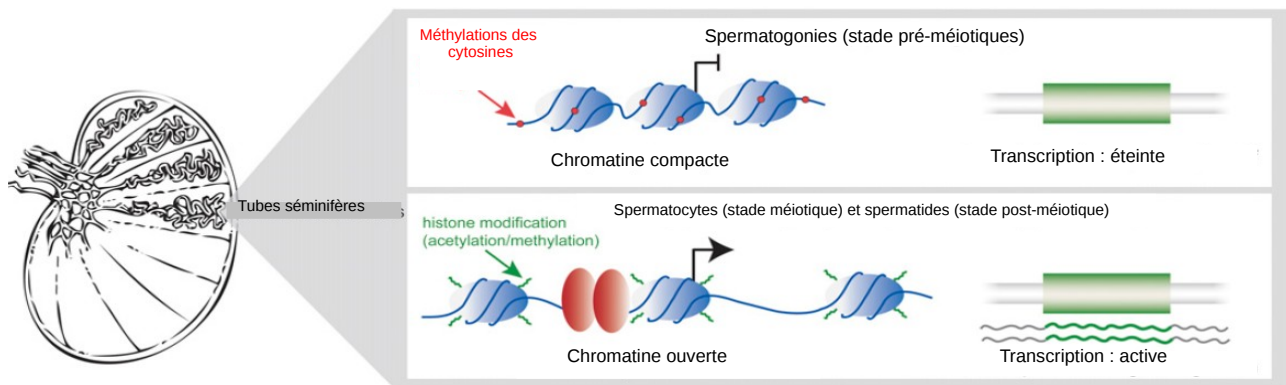


Figure 13. L'hypothèse « out of testis » pour l'émergence de nouveaux gènes

Cette hypothèse suggère que la transcription de nouvelles copies géniques est facilitée dans certaines cellules germinales des testicules (aux stades méiotiques et post méiotiques), car à ces stades, la chromatine est dite globalement permissive (c'est à dire globalement ouverte) et certains composants de la machinerie de transcription sont particulièrement surexprimés. Chez les mammifères, la déméthylation globale des îlots CpG au niveau des séquences promotrices des gènes et les modifications d'histones (ronds bleus), faciliteraient l'accès aux composants de la machinerie transcriptionnelle (ronds rouges). Une fois transcrits, les nouveaux gènes fonctionnels (petites vagues vertes), qui présentent un avantage pour l'hôte, peuvent être sélectionnés et à termes, acquérir des promoteurs plus efficaces (ce processus est facilité dans les testicules car l'expression spécifiques de gènes dans ces tissus, nécessite des promoteurs comportant des structures simples). Sur le long terme, certains nouveaux gènes peuvent éventuellement évoluer vers des profils d'expression plus diversifiés et donc acquérir de nouvelles fonctions cellulaires dans d'autres tissus. Adapté de (Kaessmann, 2010).

Conclusion

L'ensemble des résultats mettent en perspectives la dynamique des ET ainsi que celle apportée par le génome *via* la machinerie des *piARN*.

Au niveau de la lignée M19, nous avons montré que la dynamique de l'élément *mos1* était stabilisée et associée à sa répression transcriptionnelle dans la lignée germinale. Néanmoins, sa répression par la voie des *piARN* ne semble pas être associée à l'absence de mobilisation de l'élément. En effet, les insertions caractérisées traduisent un polymorphisme d'insertion élevé suggérant que *mos1* soit régulièrement mobilisé. Cette observation contraste fortement avec l'apparente stabilité en termes de nombre de copies et indique que la voie des *piARN* aide à limiter l'expansion des nouveaux ET émergents. Au vu de ces résultats, il semble que la voie des *piARN* favorise l'établissement d'un état d'équilibre entre le génome et ses parasites.

Au niveau populationnel, l'élément *mos1* est caractérisé par ses variations d'activité, et son hétérogénéité entre les différentes lignées de génome. Nous avons pu associer ces variations au fort polymorphisme transcriptionnel des gènes associés à la régulation des *piARN*. Ce polymorphisme rend compte d'un fort dimorphisme sexuel dans lequel *mos1* est fortement surexprimé dans les testicules tandis que les gènes de la voie des *piARN* y sont très nettement sous-exprimés. Par ailleurs, les niveaux d'expression de *mos1* sont négativement corrélés à ceux d'*ago3* dans les ovaires des populations colonisatrices uniquement, suggérant un lien entre l'efficacité de la machinerie des *piARN* et l'activité de *mos1*. Le scénario évolutif présenté en discussion de ce travail est adapté à *mos1*. J'espère qu'il donnera suite à d'autres travaux, d'une part pour confirmer le cas de *mos1*, d'autre part, pour intégrer dans un modèle général, l'activité des ET, le polymorphisme (de séquence, transcriptionnel, traductionnel) de l'ensemble des gènes associés à la régulation des *piARN*, les variations de production de *piARN*, ainsi que l'histoire démographique associée au processus de colonisation des espèces (e.g. *D. melanogaster* et *D. simulans*).

Au niveau interspécifique, nous avons observé que le contenu en ET chez les espèces sœurs *D. melanogaster* et *D. simulans* avait fortement divergé et par ailleurs, que cela avait affecté la réponse du génome induite par la voie des *piARN*. Chez *D. melanogaster*, de nombreux « bursts » de transposition ont eu lieu récemment. Ces familles d'ET sont activement réprimées par les *piARN* dans l'ovaire et donc, se retrouvent massivement surexprimés dans les testicules. Chez *D. simulans*, la plupart des ET sont fragmentés et résultent probablement de vagues d'amplifications plus anciennes. Nous supposons donc que la réponse par les *piARN* chez *D. simulans* résulte principalement d'une régulation passée qui semble être la relique d'anciennes invasions d'ET.

Nous avons enfin pu constater des différences majeures entre les lignées germinales mâles et femelles au cours de ces trois études. Premièrement, l'activité transcriptionnelle des ET semble plus élevée dans les testicules en comparaison aux ovaires. Deuxièmement, les gènes associés à la voie des *piARN* dans la lignée germinale sont surexprimés dans les ovaires. Troisièmement, la production de *piARN* est nettement moins

efficace dans les testicules que dans les ovaires. Nous supposons que ces différences sont la conséquence de pressions de sélection femelles spécifiques : d'une part pour lutter contre la transmission verticale des ET, d'autre part, pour optimiser la survie de l'embryon *via* le matériel cytoplasmique transmis maternellement. Chez les mâles, au contraire, il semble que l'activation des ET s'inscrit dans un contexte favorable à l'émergence de nouveautés génétiques. J'espère que ces hypothèses seront testées par la suite et notamment, que l'on puisse quantifier de manière exhaustive l'implication des ET dans la course à l'adaptation et à l'innovation *via* leur activité dans la lignée germinale mâle.

Bibliographie

- Abe, H., T. Fujii, T. Shimada, and K. Mita. 2010. Novel non-autonomous transposable elements on W chromosome of the silkworm, *Bombyx mori*. *J Genet.* 89:375-387.
- Abrusan, G., J. Giordano, and P.E. Warburton. 2008. Analysis of transposon interruptions suggests selection for L1 elements on the X chromosome. *PLoS Genet.* 4:e1000172.
- Alfoldi, J., F. Di Palma, M. Grabherr, C. Williams, L. Kong, E. Mauceli, P. Russell, C.B. Lowe, R.E. Glor, J.D. Jaffe, D.A. Ray, S. Boissinot, A.M. Shedlock, C. Botka, T.A. Castoe, J.K. Colbourne, M.K. Fujita, R.G. Moreno, B.F. ten Hallers, D. Haussler, A. Heger, D. Heiman, D.E. Janes, J. Johnson, P.J. de Jong, M.Y. Koriabine, M. Lara, P.A. Novick, C.L. Organ, S.E. Peach, S. Poe, D.D. Pollock, K. de Queiroz, T. Sanger, S. Searle, J.D. Smith, Z. Smith, R. Swofford, J. Turner-Maier, J. Wade, S. Young, A. Zadissa, S.V. Edwards, T.C. Glenn, C.J. Schneider, J.B. Losos, E.S. Lander, M. Breen, C.P. Ponting, and K. Lindblad-Toh. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature.* 477:587-591.
- Anne, J., and B.M. Mechler. 2005. Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuleen. *Development.* 132:2167-2177.
- Anxolabehere, D., M.G. Kidwell, and G. Periquet. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Mol Biol Evol.* 5:252-269.
- Arabidopsis Genome, I. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature.* 408:796-815.
- Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M.J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J.J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, and T. Tuschl. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature.* 442:203-207.
- Aravin, A.A., and G.J. Hannon. 2008. Small RNA silencing pathways in germ and stem cells. *Cold Spring Harb Symp Quant Biol.* 73:283-290.
- Aravin, A.A., G.J. Hannon, and J. Brennecke. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science.* 318:761-764.
- Aravin, A.A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K.F. Toth, T. Bestor, and G.J. Hannon. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell.* 31:785-799.
- Aravin, A.A., G.W. van der Heijden, J. Castaneda, V.V. Vagin, G.J. Hannon, and A. Bortvin. 2009. Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet.* 5:e1000764.
- Arkhipova, I., and M. Meselson. 2000. Transposable elements in sexual and ancient asexual taxa. *Proc Natl Acad Sci U S A.* 97:14473-14477.
- Artieri, C.G., and R.S. Singh. 2010. Molecular evidence for increased regulatory conservation during metamorphosis, and against deleterious cascading effects of hybrid breakdown in *Drosophila*. *BMC Biol.* 8:26.
- Awe, S., and R. Renkawitz-Pohl. 2010. Histone H4 acetylation is essential to proceed from a histone- to a protamine-based chromatin structure in spermatid nuclei of *Drosophila melanogaster*. *Syst Biol Reprod Med.* 56:44-61.
- Ayroles, J.F., M.A. Carbone, E.A. Stone, K.W. Jordan, R.F. Lyman, M.M. Magwire, S.M. Rollmann, L.H. Duncan, F. Lawrence, R.R. Anholt, and T.F. Mackay. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet.* 41:299-307.
- Bachtrog, D. 2003. Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol Biol Evol.* 20:173-181.
- Bachtrog, D. 2008. Evidence for male-driven evolution in *Drosophila*. *Mol Biol Evol.* 25:617-619.
- Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet.* 14:113-124.
- Barbash, D.A., J. Roote, and M. Ashburner. 2000. The *Drosophila melanogaster* hybrid male rescue gene causes inviability in male and female species hybrids. *Genetics.* 154:1747-1771.
- Barbash, D.A., D.F. Siino, A.M. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc Natl Acad Sci U S A.* 100:5302-5307.
- Barbulescu, M., G. Turner, M.I. Seaman, A.S. Deinard, K.K. Kidd, and J. Lenz. 1999. Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. *Curr Biol.* 9:861-868.

- Bayes, J.J., and H.S. Malik. 2009. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science*. 326:1538-1541.
- Begun, D.J., H.A. Lindfors, A.D. Kern, and C.D. Jones. 2007. Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba*/*Drosophila erecta* clade. *Genetics*. 176:1131-1137.
- Bellott, D.W., H. Skaletsky, T. Pyntikova, E.R. Mardis, T. Graves, C. Kremitzki, L.G. Brown, S. Rozen, W.C. Warren, R.K. Wilson, and D.C. Page. 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*. 466:612-616.
- Bentley, S.D., and J. Parkhill. 2004. Comparative genomic structure of prokaryotes. *Annu Rev Genet*. 38:771-792.
- Berry, B., S. Deddouche, D. Kirschner, J.L. Imler, and C. Antoniewski. 2009. Viral suppressors of RNA silencing hinder exogenous and endogenous small RNA pathways in *Drosophila*. *PLoS One*. 4:e5866.
- Betran, E., K. Thornton, and M. Long. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res*. 12:1854-1859.
- Bickham, J.W. 1981. Two-hundred-million-year-old chromosomes: deceleration of the rate of karyotypic evolution in turtles. *Science*. 212:1291-1293.
- Bingham, P.M., M.G. Kidwell, and G.M. Rubin. 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell*. 29:995-1004.
- Blass, E., M. Bell, and S. Boissinot. 2012. Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback. *Genome Biol Evol*. 4:687-702.
- Blumenstiel, J.P. 2011. Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet*. 27:23-31.
- Boutin, T.S., A. Le Rouzic, and P. Capy. 2012. How does selfing affect the dynamics of selfish transposable elements? *Mob DNA*. 3:5.
- Branciforte, D., and S.L. Martin. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol*. 14:2584-2592.
- Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 128:1089-1103.
- Brideau, N.J., H.A. Flores, J. Wang, S. Maheshwari, X. Wang, and D.A. Barbash. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*. 314:1292-1295.
- Bringaud, F., E. Ghedin, N.M. El-Sayed, and B. Papadopoulos. 2008. Role of transposable elements in trypanosomatids. *Microbes Infect*. 10:575-581.
- Brink, R.A., and R.A. Nilan. 1952. The Relation between Light Variegated and Medium Variegated Pericarp in Maize. *Genetics*. 37:519-544.
- Brookfield, J.F., E. Montgomery, and C.H. Langley. 1984. Apparent absence of transposable elements related to the P elements of *D. melanogaster* in other species of *Drosophila*. *Nature*. 310:330-332.
- Bucheton, A. 1973. [Study of non Mendelian female sterility in *Drosophila melanogaster*. Hereditary transmission of the degree of efficacy of the reactor factor]. *C R Acad Sci Hebd Seances Acad Sci D*. 276:641-644.
- Cabrero, J., R.J. Palomino-Morales, and J.P. Camacho. 2007. The DNA-repair Ku70 protein is located in the nucleus and tail of elongating spermatids in grasshoppers. *Chromosome Res*. 15:1093-1100.
- Campbell, A. 1962. Episomes. *Advances in Genetics*. 11:101-145.
- Cantrell, M.A., L. Scott, C.J. Brown, A.R. Martinez, and H.A. Wichman. 2008. Loss of LINE-1 activity in the megabats. *Genetics*. 178:393-404.
- Cantrell, S.R., and W.A. Bresnahan. 2005. Interaction between the human cytomegalovirus UL82 gene product (pp71) and hDaxx regulates immediate-early gene expression and viral replication. *J Virol*. 79:7792-7802.
- Carlton, J.M., R.P. Hirt, J.C. Silva, A.L. Delcher, M. Schatz, Q. Zhao, J.R. Wortman, S.L. Bidwell, U.C. Alsmark, S. Besteiro, T. Sicheritz-Ponten, C.J. Noel, J.B. Dacks, P.G. Foster, C. Simillion, Y. Van de Peer, D. Miranda-Saavedra, G.J. Barton, G.D. Westrop, S. Muller, D. Dessi, P.L. Fiori, Q. Ren, I. Paulsen, H. Zhang, F.D. Bastida-Corcuera, A. Simoes-Barbosa, M.T. Brown, R.D. Hayes, M. Mukherjee, C.Y. Okumura, R. Schneider, A.J. Smith, S. Vanacova, M. Villalvazo, B.J. Haas, M. Pertea, T.V. Feldblyum, T.R. Utterback, C.L. Shu, K. Osoegawa, P.J. de Jong, I. Hrdy, L. Horvathova, Z. Zubacova, P. Dolezal, S.B. Malik, J.M. Logsdon, Jr., K. Henze, A. Gupta, C.C. Wang, R.L. Dunne, J.A. Upcroft, P. Upcroft, O. White, S.L. Salzberg, P. Tang, C.H. Chiu, Y.S. Lee, T.M. Embley, G.H. Coombs,

- J.C. Mottram, J. Tachezy, C.M. Fraser-Liggett, and P.J. Johnson. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*. 315:207-212.
- Carrel, L., C. Park, S. Tyekucheva, J. Dunn, F. Chiaromonte, and K.D. Makova. 2006. Genomic environment predicts expression patterns on the human inactive X chromosome. *PLoS Genet*. 2:e151.
- Carvalho, A.B. 2002. Origin and evolution of the *Drosophila* Y chromosome. *Curr Opin Genet Dev*. 12:664-668.
- Carvalho, A.B., and A.G. Clark. 2005. Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila* Y. *Science*. 307:108-110.
- Carvalho, A.B., B.A. Dobo, M.D. Vibranovski, and A.G. Clark. 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 98:13225-13230.
- Casavant, N.C., L. Scott, M.A. Cantrell, L.E. Wiggins, R.J. Baker, and H.A. Wichman. 2000. The end of the LINE?: lack of recent L1 activity in a group of South American rodents. *Genetics*. 154:1809-1817.
- Cenci, G., L. Ciapponi, and M. Gatti. 2005. The mechanism of telomere protection: a comparison between *Drosophila* and humans. *Chromosoma*. 114:135-145.
- Chain, A.C., S. Zollman, J.C. Tseng, and F.A. Laski. 1991. Identification of a cis-acting sequence required for germ line-specific splicing of the P element ORF2-ORF3 intron. *Mol Cell Biol*. 11:1538-1546.
- Chang, B.H., L.C. Shimmin, S.K. Shyue, D. Hewett-Emmett, and W.H. Li. 1994. Weak male-driven molecular evolution in rodents. *Proc Natl Acad Sci U S A*. 91:827-831.
- Charlesworth, B., and C.H. Langley. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics*. 112:359-383.
- Charlesworth, B., and C.H. Langley. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet*. 23:251-287.
- Charlesworth, B., P. Sniegowski, and W. Stephan. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*. 371:215-220.
- Charlesworth, B.C., D 1983. The population dynamics of transposable elements. *Genet Res*. 42:1-27.
- Charlesworth, D., B. Charlesworth, and G. Marais. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)*. 95:118-128.
- Chen, C., J. Jin, D.A. James, M.A. Adams-Cioaba, J.G. Park, Y. Guo, E. Tenaglia, C. Xu, G. Gish, J. Min, and T. Pawson. 2009. Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc Natl Acad Sci U S A*. 106:20336-20341.
- Chippindale, A.K., and W.R. Rice. 2001. Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 98:5677-5682.
- Cioffi, M.B., and L.A. Bertollo. 2012. Chromosomal distribution and evolution of repetitive DNAs in fish. *Genome Dyn*. 7:197-221.
- Clark, J.B., W.P. Maddison, and M.G. Kidwell. 1994. Phylogenetic analysis supports horizontal transfer of P transposable elements. *Mol Biol Evol*. 11:40-50.
- Connallon, T., and L.L. Knowles. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet*. 21:495-499.
- Cullen, B.R. 2006. Viruses and microRNAs. *Nat Genet*. 38 Suppl:S25-30.
- Cutter, A.D., and S. Ward. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol Biol Evol*. 22:178-188.
- Daniels, S.B., K.R. Peterson, L.D. Strausbaugh, M.G. Kidwell, and A. Chovnick. 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*. 124:339-355.
- de Vanssay, A., A.L. Bouge, A. Boivin, C. Hermant, L. Teyssset, V. Delmarre, C. Antoniewski, and S. Ronsseray. 2012. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature*. 490:112-115.
- Deakin, J.E., T.A. Hore, E. Koina, and J.A. Marshall Graves. 2008. The status of dosage compensation in the multiple X chromosomes of the platypus. *PLoS Genet*. 4:e1000140.
- Desset, S., N. Buchon, C. Meignin, M. Coiffet, and C. Vaury. 2008. In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS One*. 3:e1526.
- Desset, S., C. Meignin, B. Dastugue, and C. Vaury. 2003. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics*. 164:501-509.
- Devlin, R.H., G.W. Stone, and D.E. Smailus. 1998. Extensive direct-tandem organization of a long repeat DNA sequence on the Y chromosome of chinook salmon (*Oncorhynchus tshawytscha*). *J Mol Evol*. 46:277-287.

- Ding, S.W. 2010. RNA-based antiviral immunity. *Nat Rev Immunol.* 10:632-644.
- Dolgin, E.S., and B. Charlesworth. 2006. The fate of transposable elements in asexual populations. *Genetics.* 174:817-827.
- Dolgin, E.S., and B. Charlesworth. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics.* 178:2169-2177.
- Dollinger, E.J. 1954. Studies on Induced Mutation in Maize. *Genetics.* 39:750-766.
- Doolittle, W.F., and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature.* 284:601-603.
- Duvernell, D.D., S.R. Pryor, and S.M. Adams. 2004. Teleost fish genomes contain a diverse array of L1 retrotransposon lineages that exhibit a low copy number and high rate of turnover. *J Mol Evol.* 59:298-308.
- Eichler, E.E., and D. Sankoff. 2003. Structural dynamics of eukaryotic chromosome evolution. *Science.* 301:793-797.
- Elde, N.C., S.J. Child, M.T. Eickbush, J.O. Kitzman, K.S. Rogers, J. Shendure, A.P. Geballe, and H.S. Malik. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell.* 150:831-841.
- Ellegren, H. 2010. Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol Evol.* 25:283-291.
- Ellegren, H., and A.K. Fridolfsson. 1997. Male-driven evolution of DNA sequences in birds. *Nat Genet.* 17:182-184.
- Ellegren, H., and A.K. Fridolfsson. 2003. Sex-specific mutation rates in salmonid fish. *J Mol Evol.* 56:458-463.
- Ellegren, H., L. Hultin-Rosenberg, B. Brunstrom, L. Dencker, K. Kultima, and B. Scholz. 2007. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol.* 5:40.
- Ellegren, H., and J. Parsch. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* 8:689-698.
- Emerson, R.A. 1914. Multiple Factors Vs. "Golden Mean" in Size Inheritance. *Science.* 40:57-58.
- Engels, W.R. 1981. Germline hypermutability in *Drosophila* and its relation to hybrid dysgenesis and cytotype. *Genetics.* 98:565-587.
- Engels, W.R., and C.R. Preston. 1979. Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility. *Genetics.* 92:161-174.
- Engels, W.R., and C.R. Preston. 1981. Identifying P factors in *Drosophila* by means of chromosome breakage hotspots. *Cell.* 26:421-428.
- Erlandsson, R., J.F. Wilson, and S. Paabo. 2000. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. *Mol Biol Evol.* 17:804-812.
- Erwin, A.A., M.A. Galdos, M.L. Wickersheim, C.C. Harrison, K.D. Marr, J.M. Colicchio, and J.P. Blumenstiel. 2015. piRNAs Are Associated with Diverse Transgenerational Effects on Gene and Transposon Expression in a Hybrid Dysgenic Syndrome of *D. virilis*. *PLoS Genet.* 11:e1005332.
- Fablet, M., A. Akkouché, V. Braman, and C. Vieira. 2014. Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis. *Gene.* 537:149-153.
- Fedoroff, N.V. 1989. About maize transposable elements and development. *Cell.* 56:181-191.
- Feschotte, C., N. Jiang, and S.R. Wessler. 2002. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet.* 3:329-341.
- Feschotte, C., and E.J. Pritham. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet.* 41:331-368.
- Fire, A., S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, and C.C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* 391:806-811.
- Forconi, M., D. Chalopin, M. Barucca, M.A. Biscotti, G. De Moro, D. Galiana, M. Gerdol, A. Pallavicini, A. Canapa, E. Olmo, and J.N. Volff. 2014. Transcriptional activity of transposable elements in coelacanth. *J Exp Zool B Mol Dev Evol.* 322:379-389.
- Furano, A.V. 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog Nucleic Acid Res Mol Biol.* 64:255-294.
- Furano, A.V., D.D. Duvernell, and S. Boissinot. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* 20:9-14.
- Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E. Nelson, S. Bowman, I.T. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S. Kyes, M.S. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft,

- M.W. Mather, A.B. Vaidya, D.M. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S. Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter, D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, and B. Barrell. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 419:498-511.
- Girard, A., R. Sachidanandam, G.J. Hannon, and M.A. Carmell. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*. 442:199-202.
- Gisselsson, D., M. Hoglund, F. Mertens, B. Johansson, P. Dal Cin, H. Van den Berghe, W.C. Earnshaw, F. Mitelman, and N. Mandahl. 1999. The structure and dynamics of ring chromosomes in human neoplastic and non-neoplastic cells. *Hum Genet*. 104:315-325.
- Good, A.G., G.A. Meister, H.W. Brock, T.A. Grigliatti, and D.A. Hickey. 1989. Rapid spread of transposable elements in experimental populations of *Drosophila melanogaster*. *Genetics*. 122:387-396.
- Good, J.M., and M.W. Nachman. 2005. Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol Biol Evol*. 22:1044-1052.
- Graveley, B.R., A.N. Brooks, J.W. Carlson, M.O. Duff, J.M. Landolin, L. Yang, C.G. Artieri, M.J. van Baren, N. Boley, B.W. Booth, J.B. Brown, L. Cherbas, C.A. Davis, A. Dobin, R. Li, W. Lin, J.H. Malone, N.R. Mattiuzzo, D. Miller, D. Sturgill, B.B. Tuch, C. Zaleski, D. Zhang, M. Blanchette, S. Dudoit, B. Eads, R.E. Green, A. Hammonds, L. Jiang, P. Kapranov, L. Langton, N. Perrimon, J.E. Sandler, K.H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P.J. Bickel, S.E. Brenner, M.R. Brent, P. Cherbas, T.R. Gingeras, R.A. Hoskins, T.C. Kaufman, B. Oliver, and S.E. Celniker. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature*. 471:473-479.
- Grégoire, M.C.L., F. ; Boissonneault, G. 2011. Sperm chromatin: biological and clinical applications in male infertility and assisted reproduction. *New York, NY: Springer*:p. 307-320.
- Gregory, T.R. 2004. Insertion-deletion biases and the evolution of genome size. *Gene*. 324:15-34.
- Griffin, D.K., L.B. Robertson, H.G. Tempest, and B.M. Skinner. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res*. 117:64-77.
- Griffin, D.K., L.B. Robertson, H.G. Tempest, A. Vignal, V. Fillon, R.P. Crooijmans, M.A. Groenen, S. Deryusheva, E. Gaginskaya, W. Carre, D. Waddington, R. Talbot, M. Volker, J.S. Masabanda, and D.W. Burt. 2008. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics*. 9:168.
- Gunawardane, L.S., K. Saito, K.M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M.C. Siomi. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*. 315:1587-1590.
- Haase, A.D., S. Fenoglio, F. Muerdter, P.M. Guzzardo, B. Czech, D.J. Pappin, C. Chen, A. Gordon, and G.J. Hannon. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev*. 24:2499-2504.
- Haldane, J.B. 1947. The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. *Ann Eugen*. 13:262-271.
- Hall, W.P. 1973a. Comparative population cytogenetics, speciation, and evolution of the iguanid lizard genus *Sceloporus*.
- Hall, W.P.S.R.K. 1973b. Hybridization of karyotypically differentiated populations in the *Sceloporus grammicus* complex (Iguanidae). *Evolution*.
- Hanson, F.B., and F.M. Heys. 1928. The Effects of Radium in Producing Lethal Mutations in *Drosophila Melanogaster*. *Science*. 68:115-116.
- Haoudi, A., M. Rachidi, M.H. Kim, S. Champion, M. Best-Belpomme, and C. Maisonhaute. 1997. Developmental expression analysis of the 1731 retrotransposon reveals an enhancement of Gag-Pol frameshifting in males of *Drosophila melanogaster*. *Gene*. 196:83-93.
- Harrison, P.W., A.E. Wright, F. Zimmer, R. Dean, S.H. Montgomery, M.A. Pointer, and J.E. Mank. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc Natl Acad Sci U S A*. 112:4393-4398.
- Hartl, D.L., A.R. Lohe, and E.R. Lozovskaya. 1997. Regulation of the transposable element mariner. *Genetica*. 100:177-184.
- Hayes, W. 1952. Recombination in *Bact. coli* K 12; unidirectional transfer of genetic material. *Nature*. 169:118-119.
- Hedges, R.W., and A.E. Jacob. 1974. Transposition of ampicillin resistance from RP4 to other replicons. *Mol Gen Genet*. 132:31-40.
- Heffron, F., C. Rubens, and S. Falkow. 1975. Translocation of a plasmid DNA sequence which mediates

- ampicillin resistance: molecular nature and specificity of insertion. *Proc Natl Acad Sci U S A*. 72:3623-3627.
- Heger, A., and C.P. Ponting. 2007. Variable strength of translational selection among 12 *Drosophila* species. *Genetics*. 177:1337-1348.
- Heim, S., N. Mandahl, U. Kristoffersson, F. Mitelman, B. Rooser, A. Rydholm, and H. Willen. 1987. Marker ring chromosome--a new cytogenetic abnormality characterizing lipogenic tumors? *Cancer Genet Cytogenet*. 24:319-326.
- Helleu, Q., P.R. Gerard, and C. Montchamp-Moreau. 2015. Sex chromosome drive. *Cold Spring Harb Perspect Biol*. 7:a017616.
- Hellsten, U., R.M. Harland, M.J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N.H. Putnam, S. Shu, L. Taher, I.L. Blitz, B. Blumberg, D.S. Dichmann, I. Dubchak, E. Amaya, J.C. Detter, R. Fletcher, D.S. Gerhard, D. Goodstein, T. Graves, I.V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S.M. Lucas, P.E. Mead, T. Mitros, H. Ogino, Y. Ohta, A.V. Poliakov, N. Pollet, J. Robert, A. Salamov, A.K. Sater, J. Schmutz, A. Terry, P.D. Vize, W.C. Warren, D. Wells, A. Wills, R.K. Wilson, L.B. Zimmerman, A.M. Zorn, R. Grainger, T. Grammer, M.K. Khokha, P.M. Richardson, and D.S. Rokhsar. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science*. 328:633-636.
- Herpin, A., S. Nakamura, T.U. Wagner, M. Tanaka, and M. Scharl. 2009. A highly conserved cis-regulatory motif directs differential gonadal synexpression of *Dmrt1* transcripts during gonad development. *Nucleic Acids Res*. 37:1510-1520.
- Hickey, D.A. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*. 101:519-531.
- Hiraizumi, Y. 1971. Spontaneous recombination in *Drosophila melanogaster* males. *Proc Natl Acad Sci U S A*. 68:268-270.
- Holmes, E.C. 2011. The evolution of endogenous viral elements. *Cell Host Microbe*. 10:368-377.
- Honeybee Genome Sequencing, C. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*. 443:931-949.
- Hood, M.E., J. Antonovics, and B. Koskella. 2004. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics*. 168:141-146.
- Houwing, S., E. Berezikov, and R.F. Ketting. 2008. Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO J*. 27:2702-2711.
- Houwing, S., L.M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D.V. Filippov, H. Blaser, E. Raz, C.B. Moens, R.H. Plasterk, G.J. Hannon, B.W. Draper, and R.F. Ketting. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*. 129:69-82.
- Hurst, L.D., and H. Ellegren. 1998. Sex biases in the mutation rate. *Trends Genet*. 14:446-452.
- Illmensee, K., and A.P. Mahowald. 1974. Transplantation of posterior polar plasm in *Drosophila*. Induction of germ cells at the anterior pole of the egg. *Proc Natl Acad Sci U S A*. 71:1016-1020.
- Ingleby, F.C., I. Flis, and E.H. Morrow. 2015. Sex-biased gene expression and sexual conflict throughout development. *Cold Spring Harb Perspect Biol*. 7:a017632.
- Innocenti, P., and E.H. Morrow. 2010a. A joint index for the intensity of sex-specific selection. *Evolution*. 64:2775-2778.
- Innocenti, P., and E.H. Morrow. 2010b. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol*. 8:e1000335.
- International Chicken Genome Sequencing, C. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 432:695-716.
- Jones, K.W., and L. Singh. 1981. Conserved repeated DNA sequences in vertebrate sex chromosomes. *Hum Genet*. 58:46-53.
- Jones, K.W., L. Singh, and R.G. Edwards. 1987. The use of probes for the Y chromosome in preimplantation embryo cells. *Hum Reprod*. 2:439-445.
- Jurka, J., O. Kohany, A. Pavlicek, V.V. Kapitonov, and M.V. Jurka. 2004. Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci U S A*. 101:1268-1272.
- Jurka, J., M. Krnjajic, V.V. Kapitonov, J.E. Stenger, and O. Kokhanyy. 2002. Active Alu elements are passed primarily through paternal germlines. *Theor Popul Biol*. 61:519-530.
- Kaessmann, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 20:1313-1326.
- Kaessmann, H., N. Vinckenbosch, and M. Long. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet*. 10:19-31.

- Kaminker, J.S., C.M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, D.A. Wheeler, S.E. Lewis, G.M. Rubin, M. Ashburner, and S.E. Celniker. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3:RESEARCH0084.
- Kawaoka, S., N. Hayashi, Y. Suzuki, H. Abe, S. Sugano, Y. Tomari, T. Shimada, and S. Katsuma. 2009. The *Bombyx* ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA*. 15:1258-1264.
- Kelleher, E.S., and D.A. Barbash. 2013. Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense. *Mol Biol Evol.* 30:1816-1829.
- Kelleher, E.S., N.B. Edelman, and D.A. Barbash. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol.* 10:e1001428.
- Khaitovich, P., I. Hellmann, W. Enard, K. Nowick, M. Leinweber, H. Franz, G. Weiss, M. Lachmann, and S. Paabo. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*. 309:1850-1854.
- Khurana, J.S., J. Wang, J. Xu, B.S. Koppetsch, T.C. Thomson, A. Nowosielska, C. Li, P.D. Zamore, Z. Weng, and W.E. Theurkauf. 2011. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell*. 147:1551-1563.
- Kidwell, M.G. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*. 115:49-63.
- Kidwell, M.G., J.F. Kidwell, and J.A. Sved. 1977. Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics*. 86:813-833.
- Kidwell, M.G., K. Kimura, and D.M. Black. 1988. Evolution of hybrid dysgenesis potential following P element contamination in *Drosophila melanogaster*. *Genetics*. 119:815-828.
- Kidwell, M.G., and D.R. Lisch. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*. 55:1-24.
- Kidwell, M.G., and J.B. Novy. 1979. Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: Sterility Resulting from Gonadal Dysgenesis in the P-M System. *Genetics*. 92:1127-1140.
- Kimmins, S., and P. Sassone-Corsi. 2005. Chromatin remodelling and epigenetic features of germ cells. *Nature*. 434:583-589.
- Kimura, M. 1979. The neutral theory of molecular evolution. *Sci Am*. 241:98-100, 102, 108 passim.
- Kimura, M. 1981. Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. *Proc Natl Acad Sci U S A*. 78:5773-5777.
- Kirino, Y., N. Kim, M. de Planell-Saguer, E. Khandros, S. Chiorean, P.S. Klein, I. Rigoutsos, T.A. Jongens, and Z. Mourelatos. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol.* 11:652-658.
- Kiuchi, T., H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai, G. Ishihara, S. Kawaoka, S. Sugano, T. Shimada, Y. Suzuki, M.G. Suzuki, and S. Katsuma. 2014. A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature*. 509:633-636.
- Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu, J.S. Khurana, F. Zhang, N. Schultz, B.S. Koppetsch, A. Nowosielska, H. Seitz, P.D. Zamore, Z. Weng, and W.E. Theurkauf. 2009. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell*. 138:1137-1149.
- Kleene, K.C. 2001. A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mech Dev*. 106:3-23.
- Koerich, L.B., X. Wang, A.G. Clark, and A.B. Carvalho. 2008. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature*. 456:949-951.
- Kofler, R., A.J. Betancourt, and C. Schlotterer. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet*. 8:e1002487.
- Kogan, G.L., A.V. Tulin, A.A. Aravin, Y.A. Abramov, A.I. Kalmykova, C. Maisonhaute, and V.A. Gvozdev. 2003. The GATE retrotransposon in *Drosophila melanogaster*: mobility in heterochromatin and aspects of its expression in germline tissues. *Mol Genet Genomics*. 269:234-242.
- Kojima, K., S. Kuramochi-Miyagawa, S. Chuma, T. Tanaka, N. Nakatsuji, T. Kimura, and T. Nakano. 2009. Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells. *Genes Cells*. 14:1155-1165.
- Kolaczkowski, B., D.N. Hupalo, and A.D. Kern. 2011. Recurrent adaptation in RNA interference genes across

- the *Drosophila* phylogeny. *Mol Biol Evol.* 28:1033-1042.
- Kong, A., M.L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S.A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W.S. Wong, G. Sigurdsson, G.B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D.F. Gudbjartsson, A. Helgason, O.T. Magnusson, U. Thorsteinsdottir, and K. Stefansson. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature.* 488:471-475.
- Koonin, E.V. 2011. Are there laws of genome evolution? *PLoS Comput Biol.* 7:e1002173.
- Koonin, E.V., and Y.I. Wolf. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36:6688-6719.
- Kordis, D., and F. Gubensek. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc Natl Acad Sci U S A.* 95:10704-10709.
- Kraemer, C., and E.R. Schmidt. 1993. The sex determining region of *Chironomus thummi* is associated with highly repetitive DNA and transposable elements. *Chromosoma.* 102:553-562.
- Kreizinger, J.D. 1960. Diepoxybutane as a Chemical Mutagen in *Zea Mays*. *Genetics.* 45:143-154.
- Ku, H.Y., and H. Lin. 2014. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. *Natl Sci Rev.* 1:205-218.
- Kuo, C.H., and H. Ochman. 2009. Deletional bias across the three domains of life. *Genome Biol Evol.* 1:145-152.
- Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, K. Takamatsu, S. Chuma, K. Kojima-Kita, Y. Shiromoto, N. Asada, A. Toyoda, A. Fujiyama, Y. Totoki, T. Shibata, T. Kimura, N. Nakatsuji, T. Noce, H. Sasaki, and T. Nakano. 2010. MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev.* 24:887-892.
- Kvikstad, E.M., and K.D. Makova. 2010. The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. *Genome Res.* 20:600-613.
- Lahn, B.T., and D.C. Page. 1999. Four evolutionary strata on the human X chromosome. *Science.* 286:964-967.
- Langley, C.H., K. Stevens, C. Cardeno, Y.C. Lee, D.R. Schrider, J.E. Pool, S.A. Langley, C. Suarez, R.B. Corbett-Detig, B. Kolaczowski, S. Fang, P.M. Nista, A.K. Holloway, A.D. Kern, C.N. Dewey, Y.S. Song, M.W. Hahn, and D.J. Begun. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics.* 192:533-598.
- Lankenau, S., V.G. Corces, and D.H. Lankenau. 1994. The *Drosophila* micropia retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. *Mol Cell Biol.* 14:1764-1775.
- Laski, F.A., D.C. Rio, and G.M. Rubin. 1986. Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell.* 44:7-19.
- Laski, F.A., and G.M. Rubin. 1989. Analysis of the cis-acting requirements for germ-line-specific splicing of the P-element ORF2-ORF3 intron. *Genes Dev.* 3:720-728.
- Lau, N.C., A.G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D.P. Bartel, and R.E. Kingston. 2006. Characterization of the piRNA complex from rat testes. *Science.* 313:363-367.
- Le Rouzic, A., and P. Capy. 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics.* 169:1033-1043.
- Le Rouzic, A., and G. Deceliere. 2005. Models of the population genetics of transposable elements. *Genet Res.* 85:171-181.
- Lechardeur, D., and G.L. Lukacs. 2006. Nucleocytoplasmic transport of plasmid DNA: a perilous journey from the cytoplasm to the nucleus. *Hum Gene Ther.* 17:882-889.
- Lederberg, J., and E.L. Tatum. 1946. Gene recombination in *Escherichia coli*. *Nature.* 158:558.
- Lee, C.C., Y.M. Mul, and D.C. Rio. 1996. The *Drosophila* P-element KP repressor protein dimerizes and interacts with multiple sites on P-element DNA. *Mol Cell Biol.* 16:5616-5622.
- Lee, Y.C. 2015. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in *Drosophila melanogaster*. *PLoS Genet.* 11:e1005269.
- Lee, Y.C., and C.H. Langley. 2012. Long-term and short-term evolutionary impacts of transposable elements on *Drosophila*. *Genetics.* 192:1411-1432.
- Levine, M.T., and H.S. Malik. 2013. A rapidly evolving genomic toolkit for *Drosophila* heterochromatin. *Fly (Austin).* 7:137-141.
- Levine, M.T., C. McCoy, D. Vermaak, Y.C. Lee, M.A. Hiatt, F.A. Matsen, and H.S. Malik. 2012. Phylogenomic analysis reveals dynamic evolutionary history of the *Drosophila* heterochromatin protein 1 (HP1) gene family. *PLoS Genet.* 8:e1002729.

- Li, C., V.V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M.D. Horwich, M. Syrzycka, B.M. Honda, E.L. Kittler, M.L. Zapp, C. Klattenhoff, N. Schulz, W.E. Theurkauf, Z. Weng, and P.D. Zamore. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*. 137:509-521.
- Liang, L., W. Diehl-Jones, and P. Lasko. 1994. Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. *Development*. 120:1201-1211.
- Lieber, M.R., and T.E. Wilson. 2010. SnapShot: Nonhomologous DNA end joining (NHEJ). *Cell*. 142:496-496 e491.
- Lim, A.K., and T. Kai. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. *Proc Natl Acad Sci U S A*. 104:6714-6719.
- Liu, H., J.Y. Wang, Y. Huang, Z. Li, W. Gong, R. Lehmann, and R.M. Xu. 2010. Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev*. 24:1876-1881.
- Liu, Z., P.H. Moore, H. Ma, C.M. Ackerman, M. Ragiba, Q. Yu, H.M. Pearl, M.S. Kim, J.W. Charlton, J.I. Stiles, F.T. Zee, A.H. Paterson, and R. Ming. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature*. 427:348-352.
- Llera-Herrera, R., A. Garcia-Gasca, C. Abreu-Goodger, A. Huvet, and A.M. Ibarra. 2013. Identification of male gametogenesis expressed genes from the scallop Nodipecten subnodosus by suppressive subtraction hybridization and pyrosequencing. *PLoS One*. 8:e73176.
- Lohe, A.R., D. De Aguiar, and D.L. Hartl. 1997. Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. *Proc Natl Acad Sci U S A*. 94:1293-1297.
- Lohe, A.R., and D.L. Hartl. 1996. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. *Mol Biol Evol*. 13:549-555.
- Long, M., and C.H. Langley. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila. *Science*. 260:91-95.
- Lorenzi, H., M. Thiagarajan, B. Haas, J. Wortman, N. Hall, and E. Caler. 2008. Genome wide survey, discovery and evolution of repetitive elements in three Entamoeba species. *BMC Genomics*. 9:595.
- Lynch, M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci U S A*. 104 Suppl 1:8597-8604.
- Malone, C.D., and G.J. Hannon. 2009. Small RNAs as guardians of the genome. *Cell*. 136:656-668.
- Marcon, L., and G. Boissonneault. 2004. Transient DNA strand breaks during mouse and human spermiogenesis new insights in stage specificity and link to chromatin remodeling. *Biol Reprod*. 70:910-918.
- Marques, A.C., I. Dupanloup, N. Vinckenbosch, A. Reymond, and H. Kaessmann. 2005. Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol*. 3:e357.
- Marques, J.T., and R.W. Carthew. 2007. A call to arms: coevolution of animal viruses and host innate immune responses. *Trends Genet*. 23:359-364.
- Martin, A., C. Troadec, A. Boualem, M. Rajab, R. Fernandez, H. Morin, M. Pitrat, C. Dogimont, and A. Bendahmane. 2009. A transposon-induced epigenetic change leads to sex determination in melon. *Nature*. 461:1135-1138.
- Martinez-Fernandez, M., L. Bernatchez, E. Rolan-Alvarez, and H. Quesada. 2010. Semi-quantitative differences in gene transcription profiles between sexes of a marine snail by a new variant of cDNA-AFLP analysis. *Mol Ecol Resour*. 10:324-330.
- Martins, M.J., C.F. Mota, and G.A. Pearson. 2013. Sex-biased gene expression in the brown alga Fucus vesiculosus. *BMC Genomics*. 14:294.
- Matsuda, E., R. Sugioka-Sugiyama, T. Mizuguchi, S. Mehta, B. Cui, and S.I. Grewal. 2011. A homolog of male sex-determining factor SRY cooperates with a transposon-derived CENP-B protein to control sex-specific directed recombination. *Proc Natl Acad Sci U S A*. 108:18754-18759.
- Maynard-Smith, J. 1958. The Theory of Evolution.
- McCarrey, J.R., and K. Thomas. 1987. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature*. 326:501-505.
- McClintock, B. 1930. A Cytological Demonstration of the Location of an Interchange between Two Non-Homologous Chromosomes of Zea Mays. *Proc Natl Acad Sci U S A*. 16:791-796.
- McClintock, B. 1932. A Correlation of Ring-Shaped Chromosomes with Variegation in Zea Mays. *Proc Natl Acad Sci U S A*. 18:677-681.
- McClintock, B. 1944. The Relation of Homozygous Deficiencies to Mutations and Allelic Series in Maize. *Genetics*. 29:478-502.
- McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A*. 36:344-355.
- Meisel, R.P., M.V. Han, and M.W. Hahn. 2009. A complex suite of forces drives gene traffic from Drosophila X

- chromosomes. *Genome Biol Evol.* 1:176-188.
- Melamed, E., and A.P. Arnold. 2007. Regional differences in dosage compensation on the chicken Z chromosome. *Genome Biol.* 8:R202.
- Menkis, A., D.J. Jacobson, T. Gustafsson, and H. Johannesson. 2008. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet.* 4:e1000030.
- Mevel-Ninio, M., A. Pelisson, J. Kinder, A.R. Campos, and A. Bucheton. 2007. The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics.* 175:1615-1624.
- Meyer-Ficca, M.L., H. Scherthan, A. Burkle, and R.G. Meyer. 2005. Poly(ADP-ribosyl)ation during chromatin remodeling steps in rat spermiogenesis. *Chromosoma.* 114:67-74.
- Mohn, F., G. Sienski, D. Handler, and J. Brennecke. 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell.* 157:1364-1379.
- Moissiard, G., and O. Voinnet. 2006. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. *Proc Natl Acad Sci U S A.* 103:19593-19598.
- Mouchiroud, D., G. D'Onofrio, B. Aissani, G. Macaya, C. Gautier, and G. Bernardi. 1991. The distribution of genes in the human genome. *Gene.* 100:181-187.
- Mouse Genome Sequencing, C., R.H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S.E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M.R. Brent, D.G. Brown, S.D. Brown, C. Bult, J. Burton, J. Butler, R.D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A.T. Chinwalla, D.M. Church, M. Clamp, C. Clee, F.S. Collins, L.L. Cook, R.R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K.D. Delehaunty, J. Deri, E.T. Dermitzakis, C. Dewey, N.J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D.M. Dunn, S.R. Eddy, L. Elnitski, R.D. Emes, P. Eswara, E. Eyas, A. Felsenfeld, G.A. Fewell, P. Flicek, K. Foley, W.N. Frankel, L.A. Fulton, R.S. Fulton, T.S. Furey, D. Gage, R.A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T.A. Graves, E.D. Green, S. Gregory, R. Guigo, M. Guyer, R.C. Hardison, D. Haussler, Y. Hayashizaki, L.W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D.B. Jaffe, L.S. Johnson, M. Jones, T.A. Jones, A. Joy, M. Kamal, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 420:520-562.
- Muller, H.J. 1925. The Regionally Differential Effect of X Rays on Crossing over in Autosomes of *Drosophila*. *Genetics.* 10:470-507.
- Muller, H.J. 1928. The Production of Mutations by X-Rays. *Proc Natl Acad Sci U S A.* 14:714-726.
- Muller, H.J. 1964. The Relation of Recombination to Mutational Advance. *Mutat Res.* 106:2-9.
- Nagao, A., T. Mituyama, H. Huang, D. Chen, M.C. Siomi, and H. Siomi. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. *RNA.* 16:2503-2515.
- Nam, K., and H. Ellegren. 2008. The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics.* 180:1131-1136.
- Nanda, I., K. Schlegelmilch, T. Haaf, M. Scharl, and M. Schmid. 2008. Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. *Cytogenet Genome Res.* 122:150-156.
- Nanda, I., J.N. Volff, S. Weis, C. Korting, A. Froschauer, M. Schmid, and M. Scharl. 2000. Amplification of a long terminal repeat-like element on the Y chromosome of the platyfish, *Xiphophorus maculatus*. *Chromosoma.* 109:173-180.
- Nellaker, C., T.M. Keane, B. Yalcin, K. Wong, A. Agam, T.G. Belgard, J. Flint, D.J. Adams, W.N. Frankel, and C.P. Ponting. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 13:R45.
- Nicolas, M., G. Marais, V. Hykelova, B. Janousek, V. Laporte, B. Vyskot, D. Mouchiroud, I. Negrutiu, D. Charlesworth, and F. Moneger. 2005. A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* 3:e4.
- Nielsen, R., C. Bustamante, A.G. Clark, S. Glanowski, T.B. Sackton, M.J. Hubisz, A. Fledel-Alon, D.M. Tanenbaum, D. Civello, T.J. White, J.S. J, M.D. Adams, and M. Cargill. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170.
- Nishida, K.M., T.N. Okada, T. Kawamura, T. Mituyama, Y. Kawamura, S. Inagaki, H. Huang, D. Chen, T. Kodama, H. Siomi, and M.C. Siomi. 2009. Functional involvement of Tudor and dPRMT5 in the

- piRNA processing pathway in *Drosophila* germlines. *EMBO J.* 28:3820-3831.
- Nishida, K.M., K. Saito, T. Mori, Y. Kawamura, T. Nagami-Okada, S. Inagaki, H. Siomi, and M.C. Siomi. 2007. Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA*. 13:1911-1922.
- O'Meally, D., H.R. Patel, R. Stiglec, S.D. Sarre, A. Georges, J.A. Marshall Graves, and T. Ezaz. 2010. Non-homologous sex chromosomes of birds and snakes share repetitive sequences. *Chromosome Res.* 18:787-800.
- Obbard, D.J., K.H. Gordon, A.H. Buck, and F.M. Jiggins. 2009a. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci.* 364:99-115.
- Obbard, D.J., F.M. Jiggins, D.L. Halligan, and T.J. Little. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol.* 16:580-585.
- Obbard, D.J., J.J. Welch, K.W. Kim, and F.M. Jiggins. 2009b. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5:e1000698.
- Ohta, T., and M. Kimura. 1981. Some calculations on the amount of selfish DNA. *Proc Natl Acad Sci U S A.* 78:1129-1132.
- Oliveira de Carvalho, M., J.C. Silva, and E.L. Loreto. 2004. Analyses of P-like transposable element sequences from the genome of *Anopheles gambiae*. *Insect Mol Biol.* 13:55-63.
- Olivieri, D., M.M. Sykora, R. Sachidanandam, K. Mechtler, and J. Brennecke. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J.* 29:3301-3317.
- Orgel, L.E., and F.H. Crick. 1980. Selfish DNA: the ultimate parasite. *Nature.* 284:604-607.
- Osanai, M., K.K. Kojima, R. Futahashi, S. Yaguchi, and H. Fujiwara. 2006. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene.* 376:281-289.
- Ozouf-Costaz, C., E. Pisano, C. Bonillo, and R. Williams. 1996. Ribosomal RNA location in the Antarctic fish *Champsocephalus gunnari* (Notothenioidei, Channichthyidae) using banding and fluorescence in situ hybridization. *Chromosome Res.* 4:557-561.
- Pane, A., K. Wehr, and T. Schupbach. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell.* 12:851-862.
- Pardo-Manuel de Villena, F., and C. Sapienza. 2001. Recombination is proportional to the number of chromosome arms in mammals. *Mamm Genome.* 12:318-322.
- Parisi, M., R. Nuttall, D. Naiman, G. Bouffard, J. Malley, J. Andrews, S. Eastman, and B. Oliver. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science.* 299:697-700.
- Parker, G.A., R.R. Baker, and V.G. Smith. 1972. The origin and evolution of gamete dimorphism and the male-female phenomenon. *J Theor Biol.* 36:529-553.
- Parsch, J., and H. Ellegren. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14:83-87.
- Pasyukova, E., S. Nuzhdin, W. Li, and A.J. Flavell. 1997. Germ line transposition of the copia retrotransposon in *Drosophila melanogaster* is restricted to males by tissue-specific control of copia RNA levels. *Mol Gen Genet.* 255:115-124.
- Peng, J.C., and G.H. Karpen. 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol.* 9:25-35.
- Petrov, D.A., A.S. Fiston-Lavier, M. Lipatov, K. Lenkov, and J. Gonzalez. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol.* 28:1633-1644.
- Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Sanyal, H. Kim, K. Collura, D.S. Brar, S. Jackson, R.A. Wing, and O. Panaud. 2006. Doubling genome size without polyploidization: dynamics of retrotransposon-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* 16:1262-1269.
- Pokorna, M., L. Kratochvil, and E. Kejnovsky. 2011. Microsatellite distribution on sex chromosomes at different stages of heteromorphism and heterochromatinization in two lizard species (Squamata: Eublepharidae: *Coleonyx elegans* and *lacertidae: Eremias velox*). *BMC Genet.* 12:90.
- Pokorna, M., M. Rabova, P. Rab, M.A. Ferguson-Smith, W. Rens, and L. Kratochvil. 2010. Differentiation of sex chromosomes and karyotypic evolution in the eye-lid geckos (Squamata: Gekkota: Eublepharidae), a group with different modes of sex determination. *Chromosome Res.* 18:809-820.
- Pollard, H., G. Toumaniantz, J.L. Amos, H. Avet-Loiseau, G. Guihard, J.P. Behr, and D. Escande. 2001. Ca²⁺-sensitive cytosolic nucleases prevent efficient delivery to the nucleus of injected plasmids. *J Gene Med.* 3:153-164.

- Priyadarshini, P., B.S. Murthy, J. Nagaraju, and L. Singh. 2003. A GATA-binding protein expressed predominantly in the pupal ovary of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol.* 33:185-195.
- Prud'homme, N., M. Gans, M. Masson, C. Terzian, and A. Bucheton. 1995. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics.* 139:697-711.
- Qi, H., T. Watanabe, H.Y. Ku, N. Liu, M. Zhong, and H. Lin. 2011. The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J Biol Chem.* 286:3789-3797.
- Ranz, J.M., C.I. Castillo-Davis, C.D. Meiklejohn, and D.L. Hartl. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science.* 300:1742-1745.
- Rathke, C., W.M. Baarends, S. Jayaramaiah-Raja, M. Bartkuhn, R. Renkawitz, and R. Renkawitz-Pohl. 2007. Transition from a nucleosome-based to a protamine-based chromatin configuration during spermiogenesis in *Drosophila*. *J Cell Sci.* 120:1689-1700.
- Ray, D.A., C. Feschotte, H.J. Pagan, J.D. Smith, E.J. Pritham, P. Arensburger, P.W. Atkinson, and N.L. Craig. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* 18:717-728.
- Reinke, V., I.S. Gil, S. Ward, and K. Kazmer. 2004. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development.* 131:311-323.
- Reuter, M., S. Chuma, T. Tanaka, T. Franz, A. Stark, and R.S. Pillai. 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol.* 16:639-646.
- Rhoades, M.M. 1938. Effect of the Dt Gene on the Mutability of the a(1) Allele in Maize. *Genetics.* 23:377-397.
- Ribet, D., F. Harper, A. Dupressoir, M. Dewannieux, G. Pierron, and T. Heidmann. 2008. An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* 18:597-609.
- Rice, W.R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature.* 381:232-234.
- Rice, W.R. 1998. Male fitness increases when females are eliminated from gene pool: implications for the Y chromosome. *Proc Natl Acad Sci U S A.* 95:6217-6221.
- Rice, W.R., A.D. Stewart, E.H. Morrow, J.E. Linder, N. Orteiza, and P.G. Byrne. 2006. Assessing sexual conflict in the *Drosophila melanogaster* laboratory model system. *Philos Trans R Soc Lond B Biol Sci.* 361:287-299.
- Richards, S., Y. Liu, B.R. Bettencourt, P. Hradecky, S. Letovsky, R. Nielsen, K. Thornton, M.J. Hubisz, R. Chen, R.P. Meisel, O. Couronne, S. Hua, M.A. Smith, P. Zhang, J. Liu, H.J. Bussemaker, M.F. van Batenburg, S.L. Howells, S.E. Scherer, E. Sodergren, B.B. Matthews, M.A. Crosby, A.J. Schroeder, D. Ortiz-Barrientos, C.M. Rives, M.L. Metzker, D.M. Muzny, G. Scott, D. Steffen, D.A. Wheeler, K.C. Worley, P. Havlak, K.J. Durbin, A. Egan, R. Gill, J. Hume, M.B. Morgan, G. Miner, C. Hamilton, Y. Huang, L. Waldron, D. Verduzco, K.P. Clerc-Blankenburg, I. Dubchak, M.A. Noor, W. Anderson, K.P. White, A.G. Clark, S.W. Schaeffer, W. Gelbart, G.M. Weinstock, and R.A. Gibbs. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* 15:1-18.
- Rio, D.C., and G.M. Rubin. 1988. Identification and purification of a *Drosophila* protein that binds to the terminal 31-base-pair inverted repeats of the P transposable element. *Proc Natl Acad Sci U S A.* 85:8929-8933.
- Robine, N., N.C. Lau, S. Balla, Z. Jin, K. Okamura, S. Kuramochi-Miyagawa, M.D. Blower, and E.C. Lai. 2009. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol.* 19:2066-2076.
- Romero, D., and R. Palacios. 1997. Gene amplification and genomic plasticity in prokaryotes. *Annu Rev Genet.* 31:91-111.
- Rong, Y.S. 2008. Telomere capping in *Drosophila*: dealing with chromosome ends that most resemble DNA breaks. *Chromosoma.* 117:235-242.
- Ronsseray, S., M. Lehmann, and D. Anxolabehere. 1991. The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics.* 129:501-512.
- Ronsseray, S., M. Lehmann, D. Nouaud, and D. Anxolabehere. 1996. The regulatory properties of autonomous subtelomeric P elements are sensitive to a Suppressor of variegation in *Drosophila melanogaster*.

- Genetics*. 143:1663-1674.
- Roos, C., J. Schmitz, and H. Zischler. 2004. Primate jumping genes elucidate strepsirrhine phylogeny. *Proc Natl Acad Sci U S A*. 101:10650-10654.
- Rubin, G.M., M.G. Kidwell, and P.M. Bingham. 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell*. 29:987-994.
- Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka, and S. Yuki. 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature*. 312:659-661.
- Saito, K., S. Inagaki, T. Mituyama, Y. Kawamura, Y. Ono, E. Sakota, H. Kotani, K. Asai, H. Siomi, and M.C. Siomi. 2009. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature*. 461:1296-1299.
- Saito, K., K.M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, and M.C. Siomi. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev*. 20:2214-2222.
- Saito, K., and M.C. Siomi. 2010. Small RNA-mediated quiescence of transposable elements in animals. *Dev Cell*. 19:687-697.
- Sakkas, D., and J.G. Alvarez. 2010. Sperm DNA fragmentation: mechanisms of origin, impact on reproductive outcome, and analysis. *Fertil Steril*. 93:1027-1036.
- Santos, E.M., P. Kille, V.L. Workman, G.C. Paull, and C.R. Tyler. 2008. Sexually dimorphic gene expression in the brains of mature zebrafish. *Comp Biochem Physiol A Mol Integr Physiol*. 149:314-324.
- Sassaman, D.M., B.A. Dombroski, J.V. Moran, M.L. Kimberland, T.P. Naas, R.J. DeBerardinis, A. Gabriel, G.D. Swergold, and H.H. Kazazian, Jr. 1997. Many human L1 elements are capable of retrotransposition. *Nat Genet*. 16:37-43.
- Sassone-Corsi, P. 2002. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science*. 296:2176-2178.
- Satyaki, P.R., T.N. Cuykendall, K.H. Wei, N.J. Brideau, H. Kwak, S. Aruna, P.M. Ferree, S. Ji, and D.A. Barbash. 2014. The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. *PLoS Genet*. 10:e1004240.
- Schaack, S., C. Gilbert, and C. Feschotte. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 25:537-546.
- Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, P. Van Buren, M.W. Vaughn, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 326:1112-1115.
- Schultz, N., F.K. Hamra, and D.L. Garbers. 2003. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A*. 100:12201-12206.
- Senti, K.A., and J. Brennecke. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet*. 26:499-509.
- Shapiro, J.A. 1969. Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *J Mol Biol*. 40:93-105.
- Shapiro, J.A. 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci U S A*. 76:1933-1937.
- Shimizu, N., F. Kamezaki, and S. Shigematsu. 2005. Tracking of microinjected DNA in live cells reveals the intracellular behavior and elimination of extrachromosomal genetic material. *Nucleic Acids Res*. 33:6296-6307.
- Shoji, M., T. Tanaka, M. Hosokawa, M. Reuter, A. Stark, Y. Kato, G. Kondoh, K. Okawa, T. Chujo, T. Suzuki,

- K. Hata, S.L. Martin, T. Noce, S. Kuramochi-Miyagawa, T. Nakano, H. Sasaki, R.S. Pillai, N. Nakatsuji, and S. Chuma. 2009. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell*. 17:775-787.
- Shpiz, S., S. Ryazansky, I. Olovnikov, Y. Abramov, and A. Kalmykova. 2014. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. *PLoS Genet*. 10:e1004138.
- Siebel, C.W., and D.C. Rio. 1990. Regulated splicing of the Drosophila P transposable element third intron in vitro: somatic repression. *Science*. 248:1200-1208.
- Sienski, G., D. Donertas, and J. Brennecke. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell*. 151:964-980.
- Silva, J.C., E.L. Loreto, and J.B. Clark. 2004. Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol*. 6:57-71.
- Simmons, M.J., J.D. Raymond, C.D. Grimes, C. Belinco, B.C. Haake, M. Jordan, C. Lund, T.A. Ojala, and D. Papermaster. 1996. Repression of hybrid dysgenesis in Drosophila melanogaster by heat-shock-inducible sense and antisense P-element constructs. *Genetics*. 144:1529-1544.
- Singer, M.F., V. Krek, J.P. McMillan, G.D. Swergold, and R.E. Thayer. 1993. LINE-1: a human transposable element. *Gene*. 135:183-188.
- Singh, L., and K.W. Jones. 1986. Bkm sequences are polymorphic in humans and are clustered in pericentric regions of various acrocentric chromosomes including the Y. *Hum Genet*. 73:304-308.
- Singh, L., R. Wadhwa, S. Naidu, R. Nagaraj, and M. Ganesan. 1994. Sex- and tissue-specific Bkm(GATA)-binding protein in the germ cells of heterogametic sex. *J Biol Chem*. 269:25321-25327.
- Siomi, M.C., T. Mannen, and H. Siomi. 2010. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev*. 24:636-646.
- Siomi, M.C., K. Saito, and H. Siomi. 2008. How selfish retrotransposons are silenced in Drosophila germline and somatic cells. *FEBS Lett*. 582:2473-2478.
- Siomi, M.C., K. Sato, D. Pezic, and A.A. Aravin. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol*. 12:246-258.
- Skaletsky, H., T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S.F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfsing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.P. Yang, R.H. Waterston, R.K. Wilson, S. Rozen, and D.C. Page. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 423:825-837.
- Small, C.M., G.E. Carney, Q. Mo, M. Vannucci, and A.G. Jones. 2009. A microarray analysis of sex- and gonad-biased gene expression in the zebrafish: evidence for masculinization of the transcriptome. *BMC Genomics*. 10:579.
- Snyder, M.P., D. Kimbrell, M. Hunkapiller, R. Hill, J. Fristrom, and N. Davidson. 1982. A transposable element that splits the promoter region inactivates a Drosophila cuticle protein gene. *Proc Natl Acad Sci U S A*. 79:7430-7434.
- Song, J., J. Liu, S.L. Schnakenberg, H. Ha, J. Xing, and K.C. Chen. 2014. Variation in piRNA and transposable element content in strains of Drosophila melanogaster. *Genome Biol Evol*. 6:2786-2798.
- Stadler, L.J. 1928. Mutations in Barley Induced by X-Rays and Radium. *Science*. 68:186-187.
- Stattey, J.P.O.B., S. 1998. Patterns of Y and X chromosome DNA sequence divergence during the Felidae radiation. *Genetics*.
- Steinemann, M., and S. Steinemann. 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution. *Genetica*. 102-103:409-420.
- Streck, R.D., J.E. Macgaffey, and S.K. Beckendorf. 1986. The structure of hobo transposable elements and their insertion sites. *EMBO J*. 5:3615-3623.
- Swanson, W.J., and V.D. Vacquier. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet*. 3:137-144.
- Swarts, D.C., M.M. Jore, E.R. Westra, Y. Zhu, J.H. Janssen, A.P. Snijders, Y. Wang, D.J. Patel, J. Berenguer, S.J. Brouns, and J. van der Oost. 2014a. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*. 507:258-261.
- Swarts, D.C., K. Makarova, Y. Wang, K. Nakanishi, R.F. Ketting, E.V. Koonin, D.J. Patel, and J. van der Oost.

- 2014b. The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol.* 21:743-753.
- Taylor, A.L. 1963. Bacteriophage-Induced Mutation in Escherichia Coli. *Proc Natl Acad Sci U S A.* 50:1043-1051.
- Taylor, J., S. Tyekucheva, M. Zody, F. Chiaromonte, and K.D. Makova. 2006. Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol Biol Evol.* 23:565-573.
- Torgerson, D.G., R.J. Kulathinal, and R.S. Singh. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol Biol Evol.* 19:1973-1980.
- Toro, N. 2003. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environ Microbiol.* 5:143-151.
- Touchon, M., and E.P. Rocha. 2007. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol.* 24:969-981.
- Townsend, J.P., and D.L. Hartl. 2000. The kinetics of transposable element autoregulation. *Genetica.* 108:229-237.
- Toyooka, Y., N. Tsunekawa, Y. Takahashi, Y. Matsui, M. Satoh, and T. Noce. 2000. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech Dev.* 93:139-149.
- Tribolium Genome Sequencing, C., S. Richards, R.A. Gibbs, G.M. Weinstock, S.J. Brown, R. Denell, R.W. Beeman, R. Gibbs, R.W. Beeman, S.J. Brown, G. Bucher, M. Friedrich, C.J. Grimmlikhuijzen, M. Klingler, M. Lorenzen, S. Richards, S. Roth, R. Schroder, D. Tautz, E.M. Zdobnov, D. Muzny, R.A. Gibbs, G.M. Weinstock, T. Attaway, S. Bell, C.J. Buhay, M.N. Chandrabose, D. Chavez, K.P. Clerk-Blankenburg, A. Cree, M. Dao, C. Davis, J. Chacko, H. Dinh, S. Dugan-Rocha, G. Fowler, T.T. Garner, J. Garnes, A. Gnirke, A. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S.N. Jhangiani, V. Joshi, Z.M. Khan, L. Jackson, C. Kovar, A. Kowis, S. Lee, L.R. Lewis, J. Margolis, M. Morgan, L.V. Nazareth, N. Nguyen, G. Okwuonu, D. Parker, S. Richards, S.J. Ruiz, J. Santibanez, J. Savard, S.E. Scherer, B. Schneider, E. Sodergren, D. Tautz, S. Vattahil, D. Villasana, C.S. White, R. Wright, Y. Park, R.W. Beeman, J. Lord, B. Oppert, M. Lorenzen, S. Brown, L. Wang, J. Savard, D. Tautz, S. Richards, G. Weinstock, R.A. Gibbs, Y. Liu, K. Worley, G. Weinstock, C.G. Elsik, J.T. Reese, E. Elhaik, G. Landan, D. Graur, P. Arensburger, P. Atkinson, R.W. Beeman, J. Beidler, S.J. Brown, J.P. Demuth, D.W. Drury, Y.Z. Du, H. Fujiwara, M. Lorenzen, et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature.* 452:949-955.
- Trivers, R.L. 1972. Parental investment and sexual selection. *Sexual Selection and the Descent of Man, Blackwell Science.*, p 307.
- Tucker, P.K., R.M. Adkins, and J.S. Rest. 2003. Differential rates of evolution for the ZFY-related zinc finger genes, Zfy, Zfx, and Zfa in the mouse genus Mus. *Mol Biol Evol.* 20:999-1005.
- Vagin, V.V., A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P.D. Zamore. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science.* 313:320-324.
- Vagin, V.V., J. Wohlschlegel, J. Qu, Z. Jonsson, X. Huang, S. Chuma, A. Girard, R. Sachidanandam, G.J. Hannon, and A.A. Aravin. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* 23:1749-1762.
- Vasileva, A., D. Tiedau, A. Firooznia, T. Muller-Reichert, and R. Jessberger. 2009. Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. *Curr Biol.* 19:630-639.
- Velando, A., R. Torres, and C. Alonso-Alvarez. 2008. Avoiding bad genes: oxidatively damaged DNA in germ line and mate choice. *Bioessays.* 30:1212-1219.
- Vermaak, D., S. Henikoff, and H.S. Malik. 2005. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. *PLoS Genet.* 1:96-108.
- Vermaak, D., and H.S. Malik. 2009. Multiple roles for heterochromatin protein 1 genes in Drosophila. *Annu Rev Genet.* 43:467-492.
- Vicoso, B., and D. Bachtrog. 2013. Reversal of an ancient sex chromosome to an autosome in Drosophila. *Nature.* 499:332-335.
- Vinckenbosch, N., I. Dupanloup, and H. Kaessmann. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A.* 103:3220-3225.
- Vinogradov, A.E. 1999. Intron-genome size relationship on a large evolutionary scale. *J Mol Evol.* 49:376-384.
- Vinogradov, A.E. 2005. Genome size and chromatin condensation in vertebrates. *Chromosoma.* 113:362-369.
- Volff, J.N. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb).* 94:280-294.
- Volff, J.N., L. Bouneau, C. Ozouf-Costaz, and C. Fischer. 2003. Diversity of retrotransposable elements in

- compact pufferfish genomes. *Trends Genet.* 19:674-678.
- Volff, J.N., C. Korting, and M. Scharl. 2000. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. *Mol Biol Evol.* 17:1673-1684.
- Volff, J.N., C. Korting, K. Sweeney, and M. Scharl. 1999. The non-LTR retrotransposon Rex3 from the fish *Xiphophorus* is widespread among teleosts. *Mol Biol Evol.* 16:1427-1438.
- Wagner, A. 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. *Mol Biol Evol.* 23:723-733.
- Wang, G., A. Ma, C.M. Chow, D. Horsley, N.R. Brown, I.G. Cowell, and P.B. Singh. 2000. Conservation of heterochromatin protein 1 function. *Mol Cell Biol.* 20:6970-6983.
- Ward, W.S. 2011. Regulating DNA supercoiling: sperm points the way. *Biol Reprod.* 84:841-843.
- Warren, W.C., L.W. Hillier, J.A. Marshall Graves, E. Birney, C.P. Ponting, F. Grutzner, K. Belov, W. Miller, L. Clarke, A.T. Chinwalla, S.P. Yang, A. Heger, D.P. Locke, P. Miethke, P.D. Waters, F. Veyrunes, L. Fulton, B. Fulton, T. Graves, J. Wallis, X.S. Puente, C. Lopez-Otin, G.R. Ordóñez, E.E. Eichler, L. Chen, Z. Cheng, J.E. Deakin, A. Alsop, K. Thompson, P. Kirby, A.T. Papenfuss, M.J. Wakefield, T. Olender, D. Lancet, G.A. Huttley, A.F. Smit, A. Pask, P. Temple-Smith, M.A. Batzer, J.A. Walker, M.K. Konkel, R.S. Harris, C.M. Whittington, E.S. Wong, N.J. Gemmell, E. Buschiazzi, I.M. Vargas Jentzsch, A. Merkel, J. Schmitz, A. Zemmann, G. Churakov, J.O. Kriegs, J. Brosius, E.P. Murchison, R. Sachidanandam, C. Smith, G.J. Hannon, E. Tsend-Ayush, D. McMillan, R. Attenborough, W. Rens, M. Ferguson-Smith, C.M. Lefevre, J.A. Sharp, K.R. Nicholas, D.A. Ray, M. Kube, R. Reinhardt, T.H. Pringle, J. Taylor, R.C. Jones, B. Nixon, J.L. Dacheux, H. Niwa, Y. Sekita, X. Huang, A. Stark, P. Kheradpour, M. Kellis, P. Flicek, Y. Chen, C. Webber, R. Hardison, J. Nelson, K. Hallsworth-Pepin, K. Delehaanty, C. Markovic, P. Minx, Y. Feng, C. Kremitzki, M. Mitreva, J. Glasscock, T. Wylie, P. Wohldmann, P. Thiru, M.N. Nhan, C.S. Pohl, S.M. Smith, S. Hou, M. Nefedov, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature.* 453:175-183.
- Watanabe, T., T. Nozawa, C. Aikawa, A. Amano, F. Maruyama, and I. Nakagawa. 2013. CRISPR regulation of intraspecies diversification by limiting IS transposition and intercellular recombination. *Genome Biol Evol.* 5:1099-1114.
- Wei, K.H., J.K. Grenier, D.A. Barbash, and A.G. Clark. 2014. Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 111:18793-18798.
- Weiner, A.M., P.L. Deininger, and A. Efstratiadis. 1986. Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem.* 55:631-661.
- Wilson Sayres, M.A., and K.D. Makova. 2011. Genome analyses substantiate male mutation bias in many species. *Bioessays.* 33:938-945.
- Wojtczak, A., K. Poplonska, and M. Kwiatkowska. 2008. Phosphorylation of H2AX histone as indirect evidence for double-stranded DNA breaks related to the exchange of nuclear proteins and chromatin remodeling in *Chara vulgaris* spermiogenesis. *Protoplasma.* 233:263-267.
- Woolhouse, M.E., J.P. Webster, E. Domingo, B. Charlesworth, and B.R. Levin. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet.* 32:569-577.
- Wu, D.D., X. Wang, Y. Li, L. Zeng, D.M. Irwin, and Y.P. Zhang. 2014. "Out of pollen" hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biol Evol.* 6:2822-2829.
- Xia, Q., D. Cheng, J. Duan, G. Wang, T. Cheng, X. Zha, C. Liu, P. Zhao, F. Dai, Z. Zhang, N. He, L. Zhang, and Z. Xiang. 2007. Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. *Genome Biol.* 8:R162.
- Yang, X., E.E. Schadt, S. Wang, H. Wang, A.P. Arnold, L. Ingram-Drake, T.A. Drake, and A.J. Lusis. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* 16:995-1004.
- Yannopoulos, G., N. Stamatis, M. Monastirioti, P. Hatzopoulos, and C. Louis. 1987. hobo is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. *Cell.* 49:487-495.
- Yin, H., and H. Lin. 2007. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature.* 450:304-308.
- Zambon, R.A., V.N. Vakharia, and L.P. Wu. 2006. RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. *Cell Microbiol.* 8:880-889.

- Zeh, D.W., J.A. Zeh, and Y. Ishida. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays*. 31:715-726.
- Zhang, G., C. Cowled, Z. Shi, Z. Huang, K.A. Bishop-Lilly, X. Fang, J.W. Wynne, Z. Xiong, M.L. Baker, W. Zhao, M. Tachedjian, Y. Zhu, P. Zhou, X. Jiang, J. Ng, L. Yang, L. Wu, J. Xiao, Y. Feng, Y. Chen, X. Sun, Y. Zhang, G.A. Marsh, G. Cramer, C.C. Broder, K.G. Frey, L.F. Wang, and J. Wang. 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science*. 339:456-460.
- Zhang, Z., J. Wang, N. Schultz, F. Zhang, S.S. Parhad, S. Tu, T. Vreven, P.D. Zamore, Z. Weng, and W.E. Theurkauf. 2014. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell*. 157:1353-1363.
- Zheng, K., J. Xiol, M. Reuter, S. Eckardt, N.A. Leu, K.J. McLaughlin, A. Stark, R. Sachidanandam, R.S. Pillai, and P.J. Wang. 2010. Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc Natl Acad Sci U S A*. 107:11841-11846.
- Zinder, N.D., and J. Lederberg. 1952. Genetic exchange in Salmonella. *J Bacteriol*. 64:679-699.

Titre : La régulation des éléments transposables par la voie des *pi*ARN: Les différences entre lignées germinales mâles et femelles et leurs conséquences sur la dynamique de transposition

Mots clés : *pi*ARN - *mariner* - Transposon - lignée germinale - testicules - *D. melanogaster* - évolution

Résumé: Les Eléments Transposables (ET) sont des parasites du génome caractérisés par leur capacité à se répliquer plus rapidement que les autres éléments génétiques du génome. La régulation par la voie des *pi*ARN joue un rôle essentiel pour limiter l'expansion des ET dans les lignées germinales des animaux.

La première question posée est comment le génome répond face à une nouvelle invasion par un ET. Dans ce but, nous avons introduit le transposon de Classe II *mariner* (sous-famille *mos1*) chez *D. melanogaster*, qui ne contient naturellement pas l'élément. Nous avons montré, qu'après son amplification autonome dans le génome, l'élément avait atteint un équilibre en termes de nombre de copies, depuis qu'une régulation *de novo* par les *pi*ARN avait été acquise.

Deuxièmement, nous avons étudié la mobilisation de l'élément *mariner* au cours du processus de colonisation des régions géographiques tempérées. A partir d'un large panel de populations naturelles nous avons trouvé que l'activité moyenne de *mariner* était remarquablement augmentée dans les populations non-Africaines en comparaison aux populations Africaines. Ces variations peuvent s'expliquer par un fort polymorphisme d'expression (transcriptionnel et traductionnel) des gènes de la voie des *pi*ARN.

Le troisième chapitre soutient que la forte activité des ET dans la lignée germinale mâle est un phénomène global chez les drosophiles. Par ailleurs, le contenu en ET chez les espèces sœurs (*D. melanogaster* et *D. simulans*) a fortement divergé et, cela a affecté la réponse associée à la production des *pi*ARN. Chez *D. melanogaster*, de nombreux « burst » de transposition ont eut lieu récemment. Ces familles d'ET sont activement réprimées par les *pi*ARN dans l'ovaire et donc, se retrouvent massivement surexprimés dans les testicules. Chez *D. simulans*, nous pensons que la réponse par les *pi*ARN résulte principalement d'une régulation passée qui semble être la relique d'anciennes invasions d'ET.

La voie des *pi*ARN est supposé être « garante de l'intégrité du génome » de par son rôle actif dans la défense contre les ET. Cependant, si la sélection naturelle purge les génomes de ces parasites délétères, il semble que les mécanismes de régulation de l'hôte contribuent au maintien de l'homéostasie du génome en limitant leur expansion, et quelque part en favoriser le maintien sur long terme. Ainsi, une autre interprétation pourrait être que la voie des *pi*ARN est « garante de la diversification du génome » de par son rôle à faciliter l'accumulation des ET.

Title : The regulation of transposable elements by the *pi*RNA pathway : differences between male and female germlines and consequences on transposon dynamics

Keywords : *pi*RNA - *mariner* - Transposon - germline - testes - *D. melanogaster* - evolution

Abstract: Transposable Elements (TEs) are genomic parasites characterized by their ability to replicate faster than any other genetic element in the genomes. The *pi*RNA mediated silencing is of central importance to limit TE expansion in the germline of animal species. The present dissertation explores the relationship between TEs and *pi*RNAs alongside their evolutionary dynamics.

The first question raised here was to understand how the genome responds to a new TE invasion. For that purpose, we injected a *mariner* Class II transposon into *D. melanogaster* genome that does not naturally contain the element. We found that, after its self-replication into the genome, the element have reached a copy number equilibrium since a *de novo pi*RNA mediated regulation have been acquired.

Second, we studied the *mariner* rewiring activity during the colonization of geographical temperate regions. From a large sampling of *D. simulans* natural populations, we found the mean activity of *mariner* to be strikingly higher in non-African populations compared to the African ones. These findings support the idea that selection acting on *pi*RNA effector proteins has been of central importance to explain TE lineages diversification during colonization process.

The third chapter provides evidences to propose that, the strong TE activity in testes, is a general phenomenon in *Drosophila*. We also observed that TE landscape divergence between the two sister species, have affected the genomic response mediated by the *pi*RNAs. As a response of their recent bursts of transposition, TEs overexpressed in testes are preferentially silenced by *pi*RNAs in *D. melanogaster* ovaries. By contrast, we assumed the *D. simulans pi*RNA response to be the relic of a past regulation that still persists mostly against inactive TEs.

The *pi*RNA silencing in the germline, is assumed to be the “vanguard of genome” defense and integrity due to its active role against TEs. However, while natural selection purifies the genome from its deleterious parasites, it seems that the host regulation contributes to genome homeostasis by limiting their expansion, and somehow, favors their longterm maintenance. Thus, another interpretation would have been that *pi*RNA silencing is the “vanguard of genome” diversification due to its active role in facilitating TE accumulation.